TOWARD A STANDARDIZED WAY FOR REPORTING ON ENERGY EFFICIENCY IN THE METRO AREA NETWORK

Ing. Etienne Victor Depasquale

Supervised by Prof. Ing. Saviour Zammit
Co-supervised by Prof. Franco Davoli
Prof. Raffaele Bolla

Department of Communications and Computer Engineering
University of Malta

2023

Dissertation for the degree of Doctor of Philosophy in Communications Engineering

Acknowledgments

Isn’t it a cliché to say that no man is an island? Yet how often I behave as if I were the centre of the universe! Please allow me that brief digression as my way of confessing how much I have depended, and still do, on others on this pilgrimage.

I must start by thanking my wife, Miriam, for her unwavering support. I could flood pages with ink, and still not do her justice. She has not flinched over these years of work; she has been the bedrock of my support.

My children have sacrificed time with their father, and put up with my weariness as I progressed through this work. Thank you, Max, thank you Marjan, thank you Helena.

Prof Ing Saviour Zammit: you have been a steady and supportive hand throughout. You conceal your wisdom behind your jovial and humane surface. Perhaps the latter does you even more credit than the former. I was guided well by you. Thank you, Saviour.

Prof Davoli: your immense depth of knowledge in the field was outdone only by your willingness to help and your availability at unearthly hours. I am humbled by your example. May I live up to it.

Prof Bolla: thank you for exposing me to the Green Abstraction Layer, for your invitation to join you in future work and exposure to the workings of the European Telecommunications Standards Institute.

Mark Tinka: another member of that exquisitely delightful group of people both expert and humane. I consulted you at length on practical matters and obtained insider support. I wish that all students find the same mentorship as you gave me.

Stephen Shew, Sterling Perrin, Philip Smith, Roberto Bruschi, Alan Breznick, Parth Arora … all these fine specimens of humanity volunteered to help without expecting anything in return. May you be rewarded.

Finally, I must add an apology to those who supported me and have slipped past my attention. If you read this and realize that you are one of them, please let me know and I shall post a note of thanks on LinkedIn.

AMDG.

Abstract

Energy is converted from one form to another through the activity of physical processes. The study of energy use, as it is converted from one form to another, therefore necessarily requires detailed understanding of the laws of physics that describe the behaviour of the entity responsible for the conversion (component: level 1 complexity). The complexity of the problem grows rapidly when these fundamental laws are not the ordinary means by which the behaviour of the entity is understood. This condition is common in systems: such aggregates encapsulate the behaviour of their components and obtain physical processes that are functions of the internal organization of these components (system of components: level 2 complexity). The complexity of the problem is compounded further when the ordinary means of interaction with the entity are no longer physical and material, but parametric representations of the entity's function(s). These representations might be summarized as key performance indices; a more granular knowledge of the entity's energy use may be obtained through study of the behaviour of its functions under a variety of operating conditions (multi-layered system of components: level 3 complexity). A fourth level in the hierarchy of complexity emerges with a localized system of systems; the fifth and final level of complexity is that of the geographically-dispersed system of systems. The complexity of the study of energy use by telecommunications networks falls into this fifth level.

Several problems take root in this complexity. Diversity of components; diversity of systems; diversity of architectures; laxity in terminology; diversity of players, each interested in specific roles and layers, and abuse of abstractions are just some of the highly impactful ones. These problems lead to poorly defined studies of energy use, incorrect cross-comparison of studies, weak analytical technique and over-extrapolated prognoses. It must be conceded that, notwithstanding grave limitations, these works have sown interest in the field and spurred research into better methods. Perhaps this is a common trajectory in the development of our scientific knowledge of this wonderful world.

I have primarily addressed the spatial aspect of the problem domain. Seeded by the observed laxity in architectural description and terminology, and driven by a documented failure arising out of misunderstanding of architectures, I have modelled the access portion of the metro area network in sufficient detail to support coherent analysis. Study was restricted to the metro area of the telecommunications network, as this was found to be the extent within a globally-spanning telecommunications network where fastest traffic growth was predicted. The market has been surveyed and the input gathered has been applied to validate my understanding, correct it, and to establish a firm foundation for future cycles of architecturally rigorous descriptions in support of the energy analyst.

This work develops mutual understanding between industrial and academic practitioners in two disciplines: ***sustainability in ICT***, and ***telecommunications operations***. The two groups have been approaching one another over the past ten – fifteen years, and much effort has been put in by both sides to cooperate. Sustainability researchers want to reduce telecommunications' Scope 1 (and beyond) greenhouse gas emissions; moreover, telecommunications network operators are keen to minimize the significant impact that energy use has on their operational expenditure. However, sustainability researchers have been hindered by the complexity of the object of their study, by the immaturity of methods, by the lack of methodology, and it is only recently that some consensus has emerged on good practice and the actual size of the problem (which, in the 1 – 2 % range of GHG emissions, is well short of more dire anticipations). On the other hand, while the operators are willing to share judiciously crafted questions, the detail of network architecture is not a matter of the public domain. The desire for rapprochement is there, but the modus operandi is still somewhat elusive. This work offers a contribution towards a solution of this problem.

The standardized methodology of the implementational model has been applied to map the access network, and work is in progress to describe aggregation and metro-core. The models can be integrated with the software-defined networking paradigm. Since the implementational model describes functions and locates them relative to reference points, then it can be used within controllers to interact with service functions in the data plane. The prerequisite is standardized application programming interfaces, and standardized data models that incorporate energy and/or power usage. The former role can be fulfilled by NETCONF (RFC 6241); the latter role can be fulfilled by YANG (RFC 6020), but a valid contender for the latter role is the Green Abstraction Layer (ES 203 237, ES 203 682). The Green Abstraction Layer's potential is investigated and its likelihood of adoption in the current data-plan driven exchange of link-state data is found to be poor.

Regardless of whether GAL or YANG fulfil NETCONF's content and operations layers, the energy-related notification data in the content layer cannot be generated without real-time power use models, as virtualization containers are not amenable to direct measurement of power use. The field of models is surveyed in a novel manner and contentious problems, productive approaches and significant developments are elicited.

Declaration

Plagiarism is defined as “the unacknowledged use, as one’s own, of work of another person, whether or not such work has been published, and as may be further elaborated in Faculty or University guidelines” (University Assessment Regulations, 2009, Regulation 39 (b)(i), University of Malta).

I, the undersigned, declare that the dissertation submitted is my work, except where acknowledged and referenced.

I understand that the penalties for committing a breach of the regulations include loss of marks; cancellation of examination results; enforced suspension of studies; or expulsion from the degree programme.

Course code: CCE6000

Title of work submitted: Towards a standardized way for reporting on energy efficiency in the metro area network

Date: 30th June 2023

Etienne Depasquale

Etienne-Victor Depasquale

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

100GBASE-ZR	100Gb/s baseband 80km reach
3GPP	3rd Generation Partnership Project
5G	Fifth Generation (mobile communications)
AC	Attachment Circuit
ACPI	Advanced Configuration and Power Interface
ADSL	Asymmetric Digital Subscriber Line
AF	Adaptation Function
AFNOG	AFrica Network Operators' Group
AN	Access Node (or Network, according to context)
ANSI	American National Standards Institute
API	Application Programming Interface
APOPS	Asia Pacific OPeratorS Forum
AR	Adaptive Rate
AS	Autonomous System
ATM	Asynchronous Transfer Mode
AUSNOG	AUStralian Network Operators' Group
BBF	Broadband Forum
BBU	BaseBand Unit
BGP	Border Gateway Protocol
BGP-LS	Border Gateway Protocol - Link State
BGP-LU	Border Gateway Protocol - Labeled Unicast
BMAP	Batch Markov Arrival Process
BNG	Broadband Network Gateway
B-NT	Broadband Network Termination
BRAS	Broadband Remote Access Server
BSS	Business Support System
BT	British Telecom
C-VID	Customer VLAN Identifier
CAGR	Compound Annual Growth Rate
CATV	Community Antenna TeleVision
CCAP	Converged Cable Access Platform
CCIE	Cisco Certified Internetwork Expert
CD	Compute Domain
CDN	Content Distribution (or Delivery) Network
CE	Customer Edge
CEN	Carrier Ethernet Network
CL	Convergence Layer
CLI	Convergence Layer Interface
CMCI	Cable Modem to CPE Interface
CMTS	Cable Modem Termination System
CN	Core Network
CO	Central Office
COTS	Commercial Off-The-Shelf
CPE	Customer Premises Equipment
CPRI	Common Public Radio Interface
CPU	Central Processing Unit
C-RAN	Centralized Radio Access Network
CRM	Customer Relationship Management
CSP	Communications Service Provider

CSPF	Constrained Shortest Path First
CSR	Cell-Site Router
CU	Centralized Unit
CWDM	Coarse Wavelength Division Multiplexing
D.C.	District of Columbia
DAA	Distributed Access Architecture
DAG	Directed Acyclic Graph
DCI	DataCentre Interconnect
DCSG	Disaggregated Cell-Site Gateway
DENOG	German Network Operators' Group
DFD	Data-Flow Diagram
DI	Drop-distribution Interface
DOCSIS	Data Over Cable Service Interface Specification
DP	Distribution Point
DPDK	Data Plane Development Kit
DRDRAM	Direct Rambus Dynamic Random Access Memory
DSA	Domain-Specific Architecture
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
DTE	Data Terminal Equipment
DU	Distributed Unit
DVFS	Dynamic Voltage and Frequency Scaling
DWDM	Dense Wavelength Division Multiplexing
EAE	Energy Aware Entity
EAN	Ethernet Access Node
EAS	Energy Aware State
eCPRI	evolved Common Public Radio Interface
EDFA	Erbium-Doped Fibre Amplifier
EIS	Environment Information Systems
EM	Element Manager
EMIS	Environmental Management Information Systems
eNB	evolved NodeB
ENOG	Eurasia Network Operators' Group
EPC	Evolved Packet Core
ERPS	Ethernet Ring Protection Switching
ESNOG	Spain Network Operators' Group (Grupo de Operadores de Red Españoles)
ESP	End Service Provider
ESPS	End Service Provider System
ETSI	European Telecommunications Standards Institute
EVC	Ethernet Virtual Connection
EVPN	Ethernet Virtual Private Network
FAQ	Frequently Asked Questions
FMC	Fixed-Mobile Convergence
FPGA	Field Programmable Gate Array
FRNOG	France Network Operators' Group
FRR	Fast Re-Route
FTTB	Fibre-To-The-Building/MDU/MTU
FTTH	Fibre-To-The-Home
FTTN	Fibre-To-The-Node
FWA	Fixed Wireless Access
GAL	Green Abstraction Layer

GALv2	Green Abstraction Layer version 2
GE-PON	Gigabit Ethernet Passive Optical Network
GHG	GreenHouse Gas
GII	Global Information Infrastructure
gNB	next generation NodeB
GP OS	General Purpose Operating System
GPON	Gigabit Passive Optical Network
GSI	Green Standard Interface
HCI	Human-Computer Interaction
HD	Hypervisor Domain
HFC	Hybrid Fibre Coaxial
HHP	HouseHolds Passed
HLS	Higher-layer split
HTML	HyperText Markup Language
HW	Hardware
IaDI	Intra Domain Interface
IC	Integrated Circuit
ICT	Information and Communication Technology
IDNOG	InDonesia Network Operators’ Group
IDS	Intrusion Detection System
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
IMS	IP Multimedia Subsystem
INNOG	India Network Operators’ Group
InP	Infrastructure Provider
IO	Input/Output
IOPS	Input/output Operations Per Second
IPTV	Internet Protocol TeleVision
IrDI	Inter Domain Interface
IS	Information Systems
ISDN	Integrated Services Digital Network
ISG	Industry Specification Group
IS-IS	Intermediate System – Intermediate System
ISO	International Organization for Standardization
ISP	Internet Service Provider
IT	Information Technology
ITNOG	Italian Network Operators’ Group
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union Telecommunication Standardization
IX	Internet eXchange
IXP	Internet eXchange Point
JANOG	Japan Network Operators’ Group
KVM	Kernel Virtual Machine
LACNOG	Latin American & Caribbean region Network Operators’ Group
LAN	Local Area Network
LCP	Local Control Policy
LDP	Label Distribution Protocol
LE	Local Exchange
LER	Label Edge Router
LFA	Loop-Free Alternate

LFIB	Label Forwarding Information Base
LH	Long-Haul
LIB	Label Information Base
LLC	Last-Level-Cache
LLS	Lower-layer split
LPI	Low Power Idle
LPwC	Local Power Controller
LR-PON	Long-Reach Passive Optical Network
LSA	Link State Advertisement
LSP	Label-Switched Path
LSR	Label Switching Router
LTE	Long Term Evolution
LX	Local Exchange
LXC	Linux Containers
MAN	Metro(15odelli) Area Network
MB	MegaByte
MDU	Multi-Dwelling Unit
MEC	Multi-access Edge Computing
MEF	Metro Ethernet Forum
MiB	MebiByte
MIPS	Millions of Instructions Per Second
MOS	Metal-Oxide Semiconductor
MPLS	MultiProtocol Label Switching
MSAN	Multi-Service Access Node
MSBN	Multi-Service Broadband Network
MSTP	Multiple Spanning Tree algorithm and Protocol
MTU	Multi-Tenant Unit
NANOG	North American Network Operators' Group
NAP	Network Access Point
NCP	Network Control Policy
NCTA	Internet and Television Association (formerly: National Cable Television
ND	Network Domain
NE	Network Element
NETCONF	Network Configuration Protocol
NFV	Network Function Virtualization
NFVI	Network Function Virtualization Infrastructure
NFV-MANO	Network Function Virtualization – MANagement and Orchestration
NFVO	Network Function Virtualization Orchestrator
NGN	Next Generation Networks
NG-PON	Next-Generation Passive Optical Network
NG-PON2	Next-Generation Passive Optical Network 2
NIC	Network Interface Card
NID	Network Interface Device
NLRI	Network Layer Reachability Information
NMS	Network Management System
NOG	Network Operators Group
NP	Network Provider
NSD	Network Service Descriptor
NSP	Network Service Provider
NT	Network Termination
NTT	Nippon Telephone and Telegraph

NUMA	Non-Uniform Memory Access
OADM	Optical Add-Drop Multiplexer
OAN	Optical Access Network
OASE	Open Access Seamless Evolution
OCA	Open Connect Appliance
ODN	Optical Distribution Network
OECD	Organization for Economic Co-operation and Development
OLS	Optical Line System
OLT	Optical Line Terminal
OM/OD	Optical Multiplexer / Optical Demultiplexer
ONT	Optical Network Terminal
ONU	Optical Network Unit
OPEX	Operational Expenditure
OPS	Optical Protection Switching
OS	Operating System
OSFP	Octal Small Form factor Pluggable
OSI	Open System Interconnection
OSP	OutSide Plant
OSPF	Open Shortest Path First
OSPM	Operating System directed Power Management
OSS	Operational Support System
OTN	Optical Transport Network
OTT	Over-The-Top
OvS	Open virtual Switch
P2P	Peer-to-peer
PAD	Problem-Approach-Development
PAI	Premises Attachment Interface
PB	Provider Bridging
PBB	Provider Backbone Bridging
PCALC	Path CALCulation
PCE	Path Computation Element
PCEP	Path Computation Element Protocol
PCP	Primary Connection Point
PDU	Power Distribution Unit or Protocol Data Unit, according to context
PE	Provider Edge
PON	Passive Optical Network
PP	Paid Peering
P-PsS	Performance Primitve sub-State
PsS	Primitive sub-State
PSTN	Public Switched Telephone Network
PTN	Public Telecommunications Network
PTNO	Public Telecommunications Network Operator
PTT	Postal, telegraph, and telephone service
PVR	Personal Video Recorder
QoS	Quality of Service
QSFP-DD	Quad Small Form factor Pluggable – Double Density
RAM	Random Access Memory
RAN	Radio Access Network
RAPL	Running Average Power Limit
RFC	Request For Comments
RfoG	Radio Frequency over Glass

RG	Residential Gateway
RN	Remote Node
RNC	Radio Network Controller
ROADM	Reconfigurable Optical Add-Drop Multiplexer
ROI	Return On Investment
RP	Reference Point
RPI-S	Reference Point for Interconnection – Service
RPN	Remote PHY Node
RPS	Remote PHY Shelf
RRH	Remote Radio Head
RSTP	Rapid Spanning Tree algorithm and Protocol
RSVP	ReSource reserVation Protocol
RU	Radio Unit, or Research Unit, or Rack Unit, depending on context
S-VID	Service VLAN Identifier
SAFNOG	South African Network Operators' Group
SAI	Serving Area Interface
SANOG	South Asia Network Operators' Group
SBA	Service-Based Architecture
SBI	Service-Based Interface
SDI	Serial Digital Interface
SDN	Software-Defined Networking
SDO	Standards Developing (or Development) Organization
SFC	Service Function Chain
SFP	Service Function Path
SG	Study Group
SGA	SG Analytics
SID	Segment Identifier
SLA	Service Level Agreement
SMPTE	Society of Motion Picture Television Engineers
SONET/SDH	Synchronous Optical Network/Synchronous Digital Hierarchy
SP	Service Provider
SPEC	Standard Performance Evaluation Corporation
S-PsS	Standby Primitive sub-State
SR	Segment Routing
SRH	Segment Routing Header
SR-IOV	Single-root Input/Output Virtualization
STB	Set-Top Box
STP	Spanning Tree Protocol
SWINOG	SWIss Network Operators' Group
TCP	Transmission Control Protocol
TCP/IP	Transmission Control Protocol/Internet Protocol
TDM	Time Division Multiplexing
TE	Terminal Equipment
TE	Traffic Engineering
TED	Traffic Engineering Database
TI-LFA	Topology-Independent Loop-Free Alternate
TIM	Telecom Italia Mobile
TLV	Type-Length-Value
TN	Telecommunications Network
TPC	Transaction Processing Performance Council
TTL	Time-To-Live

TV	TeleVision
UKNOF	United Kingdom Network Operators' Forum
UML	Universal Modelling Language
UNI	User-Network Interface
u-PE	User-facing Provider Edge
URL	Uniform Resource Locator
URLLC	Ultra-Reliable Low Latency Communication
US	United States
USD	United States Dollar
UUID	Universally Unique Identifier
VC	Virtualization Container
vCMTS	virtual Cable Modem Termination System
vCU	virtual Centralized Unit
VDN	Video Distribution Network
VDSL	Very high-speed Digital Subscriber Line
vDU	virtual Distributed Unit
VIM	Virtualized Infrastructure Manager
VLAN	Virtual Local Area Network
VM	Virtual Machine
VNF	Virtual Network Function
VNFC	Virtual Network Function Component
VNFM	Virtual Network Function Manager
VNI	Visual Networking Index
VoD	Video on Demand
VoIP	Voice over Internet Protocol
vOLT	virtual Optical Line Terminal
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
VPWS	Virtual Private Wire Service
vRU	virtual Radio Unit
WDM	Wavelength Division Multiplexing
WSC	Warehouse Scale Computing
xDSL	any variant of Digital Subscriber Line
XG-PON	Ten Gigabit Passive Optical Network
XGS-PON	Ten Gigabit Symmetrical Passive Optical Network
YANG	Yet Another Next Generation

LIST OF PUBLICATIONS

[1] E. V. Depasquale and C. Preist, ‘Sustainability Research in Computing: A Taxonomy’, in 29th International Conference on Environmental Informatics, EnviroInfo 2015 / 3rd International Conference on Information and Communication Technology for Sustainability, ICT4S 2015, Copenhagen, Denmark, September 7-9, 2015, Part 2: Adjunct Proceedings, V. K. Johannsen, S. Jensen, V. Wohlgemuth, C. Preist, and E. Eriksson, Eds., University of Copenhagen, 2015, pp. 239–240. [Online]. Available: http://iai-uiserv1.iai.fzk.de/ictensure/site?mod=litdb\&subject=art\&pid=X323F7061\&action=detail

[2] E.-V. Depasquale and C. Preist, ‘A case for the development of data standards for reporting projections of growth in the Internet’s energy consumption’, presented at the ICT4S, Amsterdam, Aug. 2016.

[3] R. Bolla, R. Bruschi, F. Davoli, and E.-V. Depasquale, ‘Energy-efficient management and control in video distribution networks: “legacy” hardware based solutions and perspectives of virtualized networking environments’, in Guide to Greening Video Distribution Networks – Energy-Efficient Internet Video Delivery, A. Popescu, Ed., Springer, 2018, pp. 25–57.

[4] E.-V. Depasquale, S. Zammit, F. Davoli, and R. Bolla, ‘An Implementational Model of Telecommunications Networks for the Energy Analyst’, in 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Jul. 2018, pp. 336–344. Doi: 10.1109/Cybermatics_2018.2018.00085.

[5] E.-V. Depasquale, H. A. Salam, and F. Davoli, ‘PAD: a graphical and numerical enhancement of structural coding to facilitate thematic analysis of a literature corpus’, ArXiv210713983 Cs, Jul. 2021, Accessed: Jul. 31, 2021. [Online]. Available: http://arxiv.org/abs/2107.13983

[6] E.-V. Depasquale, S. Zammit, F. Davoli, and R. Bolla, ‘The energy analyst’s need for a standard that interprets the Global Information Infrastructure in the Metro Area’, ArXiv220311994 Cs, Mar. 2022, Accessed: Mar. 25, 2022. [Online]. Available: http://arxiv.org/abs/2203.11994

[7] E.-V. Depasquale, H. A. Salam, and F. Davoli, ‘PAD: A graphical and numerical enhancement of structural coding to facilitate thematic analysis of a literature corpus’, MethodsX, vol. 9, p. 101633, Jan. 2022, doi: 10.1016/j.mex.2022.101633.

I am in the process of writing two journal papers that emerge from the work presented in chapter 8. I intend to submit one to the IEEE’s Communications Surveys and Tutorials journal, and the other to MDPI’s Sustainability journal. While the former will focus on trends and next-generation networks as foreseen by the players themselves, the latter (MDPI) will employ the G.800 modelling artefacts to the scenarios identified in chapter 8, and others which also emerge from the surveys. This will support precise identification of potential new RPs and RPI-Ns within the span of the metro area from the access node to the metro-core.

Chapter 1. Introduction

Energy efficiency in the Internet (and in computing and telecommunication networks in general) has become a significant problem, which has received increasing attention since the early years 2000 (see, e.g., [1], [2], [3] and references therein), starting from cloud computing infrastructures, and then extending to mobile and fixed networks. Various techniques can be applied for management and control of energy consumption, both at the device level (or parts thereof) and at the level of the network domain. A classification has been attempted in [1], among others. However: "if you can't measure it, you can't manage it" – so goes the tried-and-true maxim. How well can we measure energy consumption?

1.1 A chorus of concern on the state of the art in the analysis of energy consumption

"There is currently no standardized way of categorizing the components of network transmission. Therefore, the definition of system boundaries and the choice of included subsystems has been deemed the most important methodological decision". [4, Sec. 3.1]. This is the most significant challenge that emerges from the interviews held in this [4] qualitative research. This thesis directly addresses this challenge. In this introduction, further evidence is given of the validity of the underlying concern. A few anecdotes can be a useful point of departure in substantiating the claim to validity.

- Lawrence Berkeley National Laboratory's Energy Reporting Project has emphasized the importance of keeping track of which devices in buildings are consuming energy, when they do so and ***where they are located*** (my emphasis) [5].
- Masanet and Koomey [6], [7], [8] have called for precision in temporal and spatial aspects of analyses conducted on the energy- and greenhouse gas (GHG) – impacts of information and communication (ICT) infrastructure. Notably, attention is drawn to the danger of ignoring the impact of innovation, when projecting trends far into the future.
- A nascent organization – Greening of Streaming – with some noteworthy backing (e.g., Akamai, but see also [9]) has a "founding commitment … to make **accurate** public statements about energy efficiency in relation to streaming"[10]. The theme common to these apparently diverse statements is a call for accuracy in analysis of energy consumption. Their call directly invokes an understanding the physical distribution of consumers is concerned.

There is a good reason behind this call (for accuracy), and while Koomey [8] has been particularly explicit and specific about it, several other researchers have called out the challenge (see, e.g., [11] and especially [4, Sec. 3.1]) of ***the absence of standardized understanding of the ICT system infrastructure***. Here, concern is with standardized understanding of the focal point of network traffic growth: the telecommunications service provider's (CSP) network within the urban area, usually referred to (albeit not with universal consent [12]) as the ***metro area network***.

This thesis gives evidence of error that arises out of incongruent network models (Chapter 3), thereby drawing attention to the need for rigorous bases upon which to build analyses. Furthermore, it traces an evolution of network modelling through a series of steps that converge onto a format that is suited to describing current and future generations of metro area networks.

1.2 Five dimensions of the challenge to the energy analyst's task

A severe challenge faces the analyst intent on understanding the prognosis for growth of energy use in operating a telecommunications network (henceforth: the energy analyst). Five dimensions of the challenge have been found:

1. incoherent prognoses;
2. diversity in implementation;
3. diversity in the ecosystem of roles;
4. technological evolution, and
5. abstraction.

The dimensions are now described in more detail.

1. ***Incoherent prognoses:*** A coherent prognosis does not emerge from a first reading of energy use estimates; this has been dwelt on in some length in [13], where one apparent inconsistency was resolved. Moreover, I can add a contrast that supplements Koomey's anecdotes.

    a. On the one hand: in [14], it was claimed that the telecommunications network's consumption would more than double between 2009 and 2017. In the same vein, in [15, pp. 749–750], it was stated that extrapolation of energy-use data according to the growth in traffic and number of users would drive energy use from 1% (c.2008) to 10% (c. 2020) of the world's electricity supply (this extrapolation is an example of Masanet's and Koomey's 3rd pitfall [7], i.e., long-term projections).

    b. On the other hand, since 2009 [16], Ericsson and TeliaSonera have collaborated to thoroughly assess the life cycle of an ICT network. They claim [17] a reduction during a comparable period in the direct impact (energy footprint, carbon footprint) of the Swedish ICT sector and the "Entertainment and Media" sector. Their forecast for the short- to mid-term (2020) is a continuation of this trend. While their prognosis seems prima facie to drop into Masanet's and Koomey's 3rd pitfall (long-term projections), it diverts away from it through inside knowledge of a company's (Telia's) strategic drive towards replacement of copper with fiber media.

2. ***Diversity in implementation:*** Two dimensions to this diversity were observed: (a) various technologies may comprise the physical and logical form of a telecommunications network and (b) choice of implementation is strongly dependent on national (perhaps even regional) strategy.

    a. **Variety in technology:** The physical form of the network at the end closest to the mass market customer may consist of multiple access segments[1] of one or several types (PON, DSL, HFC, RAN). Each access network terminates at a collection of user-facing provider edge (uPE) devices (optical line terminal (OLT), digital subscriber loop access multiplexer (DSLAM), cable modem termination system (CMTS)). Traffic from the uPE is backhauled into a port of an aggregation node, which may be an Ethernet switch. The Ethernet (or other aggregating technology) switches in the various local exchange (LE/LX) / central office (CO) / hub sites aggregate a metropolitan area's traffic. These aggregating sites are interconnected through an optical network, which may be in a ring, horseshoe, or partial-/full-mesh physical topology. Regardless of physical topology, wavelength-division-multiplexing and reconfigurable optical add/drop multiplexers (ROADMs) may support a fully-meshed optical path set between the sites, or simply a hub-and-spoke path set with respect to a second-level aggregating set of sites.

    b. **Variety in strategy:** Direct comparisons of incumbents' energy reports are complicated by vastly different penetration rates of fiber connections. In OECD countries, the percentage of fiber connections in total broadband subscriptions [18] varies between 86.6% in Korea and 0.4% in Greece. Even within the European Union, this metric varies widely. Sweden's figure is 78% (51.7% in 2017), Germany's is 7.1% (1.6% in 2017) and Italy's is 14.2% (2.7% in 2017).

3. ***Diversity in the ecosystem of roles:*** the space of telecommunications networks hosts a diverse ecosystem of roles comprising standardization bodies, telecommunications vendors, physical infrastructure providers (PIPs), network providers (NPs), service providers (SPs), enterprise customers and residential customers. Even the capital-intensive roles like the PIPs and the NPs are richly populated, with commonly one "incumbent" and often more per national territory. This diversity leads to poor consensus in two aspects: nomenclature and segmentation.

---

[1] An explanation of the meaning of this term is delegated to Chapter 2 ("Background"). Until then, the intuitive sense of its meaning can suffice.

a. **Nomenclature**. William B. Norton unambiguously and purposefully states that "The Lexicon Is Important". The emphasis on terminology and understanding is not made in describing human anatomy but in the unfettered context of literature about the Internet. Norton observes widespread lack of understanding of terminology within the interconnection lexicon [19]. An explanation might, at least partially, lie in the lack of coherence between standards development organizations (SDOs, e.g., the International Telecommunication Union (ITU-T), Broadband Forum (BBF), Metro Ethernet Forum (MEF), CableLabs, the Institute of Electrical and Electronic Engineers (IEEE) and the Internet Engineering Task Force (IETF)) and NPs. For example, the IETF [20, Sec. 5] distinguishes between Customer-Edge equipment, Provider-Edge routers and Provider routers but another source [21, p. 34] steeped in AT&T's culture uses the terms Customer equipment, Access Routers and Backbone Routers for the same nodes in the IP layer. Another significant example is the term "customer premises equipment". On the one hand, in [20], this "is the box that a provider places with the customer. It serves two purposes: giving the customer ports to plug in to and making it possible for a provider to monitor the connectivity to the customer site." On the other hand, in [22, Fig. 8] and most CableLabs reports, this is the customer's own Information and Communication Technology (ICT) equipment, meeting the provider's cable modem (CM) or set-top box (STB) at the cable modem to CPE interface (CMCI). The high scaling factor of the customer end makes this variation in nomenclature a significant one, in terms of the impact it has on analysis of energy consumption.

b. **Segmentation**. Even if nomenclature were to be disregarded in an effort to align network models, the lack of correspondence between model parts stymies universal modelling. If this were due to technological, generational gaps between the modelled networks, this problem could be solved by organizing models by generation of technology. Unfortunately, this is not the case. The various uses of edge, aggregation and core have been dwelt at length on [23]. There, a conceptual segmentation of the Internet that may be aligned to discernible physical features, is suggested.

4. ***Change – the only constant?*** The telecommunications network is in constant evolution. I have referred to generational gaps between extant networks. The difficulty of developing a reference is compounded by the consideration of evolving technologies that are sufficiently different as to impact the physical topology. For example, long-reach passive optical networks (LR-PONs) may change the network into a radically different form that flattens the hierarchy of aggregations that characterize today's telecommunications networks. Moreover, XR Optics [24] may change the economics [25] of metro aggregation through the disgregation of classically-paired transceivers at opposing ends of an optic fiber link.

5. ***Abstraction is inherent to telecommunications, and this hides energy consumers.*** The telecommunications network (TN) consists of a stack of overlays (each of which may resolve into a set of layers) (see, for example [22, Fig. 12]), each of which provides a service that may be transparent to its overlay. The IP layer is a common denominator in TN stacks. It provides connectivity service to voice (VoIP), video (IPTV and OTT (over-the-top)) and general data services. Low latency jitter (< 50 ms) and high reliability (several nines) characterize IPTV quality of service (QoS) requirements; some emerging applications within the 5G metro area use-case-set are even more stringent. These requirements place constraints on network restoration that traditional IP routing protocols cannot meet. Requirements fulfilment is left to the IP layer's underlays, e.g., through path re-computation at the optical layer. These layers are transparent to an observer at the IP layer. Knowledge of the topology of the underlay is required to estimate the energy used in the provision of the service.

1.3 Objectives

A reader approaching the study of energy consumption in telecommunications networks will need to familiarize with two bodies of literature. One body of literature deals with the architecture of such networks. Such architecture literature describes their various segments, ranging from the first mile to the trans-oceanic backbones that link continents. It also describes current- and next-generation architectures of these segments, with emphasis on the first- and second-mile technologies. This body of literature presents the reader with the nomenclature problem (rooted in the ecosystem's diversity). The second body of literature, from academic as well as industrial sources, contributes to knowledge about the growth of telecommunications networks' energy consumption. Despite a general consensus on trends, there exist significant differences in the values published as well as in the scope of the network under study (see Chapter 3).

This basis – ecosystem diversity and data inconsistencies – leads to a core goal: to move towards a standardized way for reporting on energy efficiency in the metro area network. From this core goal, the objectives emerge as the following distinctive facets, or components:

1. identify and solve a particular instance of data inconsistency;
2. reconcile extant architectural paradigms;
3. identify successful analytical approaches to measurement of energy consumption, and
4. develop a baseline for a standardized perspective on current and future metro area networks.

1.4 Structure and contributions of this thesis

Contributions are closely aligned with the stated objectives; they also follow a trajectory through the field of study. They are presented within the following numbered list that also serves to outline this thesis's structure. Items in the list that refer to a contribution are preceded by the symbol **[C]**.

1. This introduction (Chapter 1) sets the scene, justifies the motivation for work within it and declares the objectives and contributions.
2. **[C]** Chapter 2 presents the first contribution, where a taxonomy of the broader field of study is set. The graphical component of the taxonomy is presented in Appendix 1.
3. In addition to the contribution indicated above, Chapter 2 also explains:
    a. why the process of convergence onto the metro area was chosen as the extent of scope of investigation of the telecommunications network, and
    b. the theoretical framework that demands the rigour which this thesis strives to satisfy.
4. **[C]** Chapter 3 shows how error arises out of inconsistent network models, and corrects the error through re-alignment of segment boundaries.
5. **[C]** Chapter 4 introduces the implementational model and uses it to lay the analytical foundations of this (implementational) modelling approach.
6. **[C]** Chapter 5 develops the analytical foundations of the implementational model into a form that is universal in technological and functional (i.e., functions in all layers, without abstraction) scope, up to, but excluding the access node.
7. **[C]** Chapter 6 investigates the extent of realization of the radical approach (identified and elaborated upon in Chapter 2), which is predicated upon the flexibility and agility conferred by software defined networking and network function virtualization. Systemic inculcation is sought, and realistic means of implementation are suggested. The Green Abstraction Layer (GAL), a standard for distributed power control, is investigated in the context of operational norms in computer system power control, and network traffic engineering.
8. **[C]** Chapter 6's analysis exposes GAL's dependency on accurate models of power use by virtual machines and containers. Chapter 7 is thus dedicated to surveying progress in analytical methods used by researchers to measure energy consumption. The method used to survey (described in Appendix 2, online only[2]) is itself a novelty, as it was developed to support the objective of extracting researchers' *modus operandi*, instead of focusing solely on their research's output.
9. **[C]** Chapter 8 lays groundwork for the current form of the implementational model of the metro area network. Appendix 4 (online only[3]) presents the raw data about current- and next-

---

[2] https://www.sciencedirect.com/science/article/pii/S2215016122000188

[3] https://drive.google.com/file/d/1hqoyTAUOfYpd0FgOgCKpqvbvhvHWx93I/view?usp=sharing

generation deployments, that are used to justify claims made in Chapter 8. Appendix 5 (online only[4]) carries the questionnaire used to collect the data.

10. Appendix 3 describes a high-level architecture of digital video distribution networks (VDNs). Its purpose is to support an understanding of the evolution of VDNs towards IP-centric modalities, as video is the primary driver of IP traffic growth (see sub-section 2.2.5). This topic (i.e., evolution of VDNs towards IP-centric modalities) is briefly acknowledged in sub-sub-section 4.2.2 (2), but it merits lengthier treatment than is afforded by the narrative of the main body of this thesis.

[C] In the process of establishing the foundations upon which to lay my work towards standardized reporting on energy efficiency in the metro area network, several standards, recommendations and reports produced by SDOs (Standards Development Organizations) were carefully studied. This work identified some errors, which were drawn to the attention of ITU-T Study Group (SG) 15 (G series) and SG 13 (Y series). The emails describing the errors are included in [26], [27], [28], [29], as is correspondence with the SG's rapporteur [30], affirming the validity of the observations, and the respective Corrigenda may be found in [31], [32], [33], [34].

---

[4] https://forms.gle/QtoTkhzEk4Q1BLdVA

Chapter 2. Background

2.1 A comprehensive, extensible framework to guide pursuit of a role in the development of a sustainable future

A researcher approaching the study of Sustainability may come from one of the many academic disciplines which contribute to its development. The approach will quickly reveal contributions from these many other disciplines. The pursuit of Sustainability is therefore a ***multi-disciplinary*** one, that has witnessed the formation of research communities (e.g., EnviroInfo and ICT4S) with participants from the diverse disciplinary backgrounds. Inherently, communities manifest some measure of isolation, to the general detriment of the field of study of Sustainability. Within the broad landscape of such communities, the researcher faces the genuine problem of matching his background, interests and objectives to those of the communities.

The researcher coming from the fields of Computing and Communications is no exception to this need to discern an alignment with these communities. The problem is compounded by the multi-disciplinarian organization of these two fields (Computing and Communication) themselves: for example, the Joint Curricula Task Force [35, p. 9] defines Computing as "*any goal-oriented activity requiring, benefiting from, or creating computers*" and identifies Computer Science, Computer Engineering, Information Technology, Software Development and Information Systems as separate disciplines. It may, therefore, be claimed that a researcher approaching the study of Sustainability from the background of Computing and/or Communications faces two, orthogonal dimensions of multi-disciplinarity. **Zones emerge in this two-dimensional space, with attendant labels ascribed to them**. In short order, the novice poses questions, like:

- "what is the difference between Green IT and Green Computing?", or
- "what is the difference between Green IS and Green Computing?", or
- "how does Sustainable HCI relate to any of the above?"

While these zones' labels are recognizable and liberally used, no clear and robust taxonomy of the space was found. Here, clarity and robustness can be construed through desirable derivative properties, such as:

- demarcation of the zones' boundaries (*clarity*),
- the activities in which communities therein are engaged (*clarity*),
- employment of the terminology of Sustainability; this would strengthen the taxonomist's claim of *robustness*, through organic development of the roots of the field of Sustainability.
- A graphic encapsulation of the categorization would facilitate assimilation (*clarity*).
- The taxonomy should be extensible, reflecting the ongoing process of maturation of the field.

This, then, was the first endeavour undertaken, as it appeared to be foundational to the formation of a ***Sustainability Researcher in Computing***. The product of this endeavour is a framework, themed ***Sustainability Research in Computing***, that attempts to illustrate the landscape for a novice to the field of study of Sustainability. The graphical representation of the framework is included in Appendix 1, as clear presentation requires more space than is afforded by a sheet of paper. Much of the content of the landscape has been extracted from Hilty's broad overview of the frameworks that have attempted to organise the landscape or portions thereof [36]. Green IT content and perspective was gleaned from Murugesan's and Gangadharan's work [37]; that of Sustainable HCI from DiSalvo et al.[38]. This work extends the breadth in [36] through the provision of a system for classification that substantially facilitates comparison of these frameworks and extraction of attributes of these frameworks. Some such attributes have also been identified.

The framework has the structure of a tree, rooted in the field of "Sustainability Research In Computing". The leaves, at the other extremity of the tree, are instantly recognizable focus areas, such as "Green Software", "Use of Renewable Energy Sources", "Persuasive Technology", "Geographic Information Systems", "Eco-labelling of IT Products", "Participatory Sensing", "Regulatory Compliance" and "Energy Efficient Computing". The path from root ("Sustainability Research in Computing") to leaves passes through branches that represent progressively narrower specialization. The first branching is into ***impact-oriented research*** and ***application-oriented research***. This distinction was first made in Berkhout's and Hertin's report to the OECD [39], where a classification into direct and indirect environmental impact of ICTs was made. ***Application-oriented research*** draws upon competences and knowledge from several fields; as such, elaboration of this zone of the field must be delegated to researchers from other disciplines. However, a defining observation can be offered here. This zone of the field is divided into research on Environment Information Systems (EIS) on one branch, and, on the other branch, into modelling and simulation of local and global ecosystems. In the first branch, researchers apply Information Systems and Technology to create Environmental Information Systems (EIS) that gather information about the environment. In the second branch, concern is with models about complex interactions on the global scale (e.g., climate change) and the local scale (e.g., heat transfer in buildings). Application-oriented research develops the tools and harvests the data that are then put to use, for instance, in Environmental Management Information Systems (EMIS). Environmental Management Information Systems are a helpful example of the bridge between application-oriented research and ***impact-oriented research***. Both the development and use of EMIS are directed at compliance, reporting and eco-efficiency, each of which are clearly within the notion of impact orientation. Focus areas supported by EMIS include validation of regulatory compliance, responsible disposal and recycling, power management, among others.

Impact-oriented research extends well beyond EMIS. It branches into ***Sustainable Human-Computer Interaction*** (HCI), ***Green IT/ICT*** and ***Green Information Systems*** (Green IS). These three branches in turn sub-divide in diverse ways, but the divisions are harmonized by their participation in the essence of the concept of ***order of impact***. First-order impact is the concern with sustainability of intrinsic behaviours and properties of Information and Communication Technologies (ICTs); second-order impact is the concern with the advancement of sustainability through the application of ICTs, and third-order impact is the concern with integration of endemic societal change in favour of sustainability.

Thus, "Green ***in*** IT/ICT" is the first-order-impact branch of Green IT/ICT. Green in IT/ICT deals with ***Life Cycle Assessment*** (LCA), or limits itself to phases (e.g., the use phase) thereof. It seeks to minimize material resource usage (e.g., raw material and energy) and consequences of usage (e.g., CO_2 emissions). The graphical representation of the framework (Appendix 1) shows many focus areas of "Green in IT/ICT", including "power management", "data centre design, layout and location", "green software", "green metrics, assessment tools and methodology", "server virtualization" and ***"energy efficient communication"***. ***Green in IT/ICT is, therefore, the space within which this thesis has been developed.***

"Green ***through***(by) IT/ICT" is the second-order-impact branch of the same "Green IT/ICT" origin. Research in this space investigates the ***Linked Life Cycle***. Sub-dividing further, a first-order- and a second-order-impact branch are discernable. Research in the first-order-impact branch approaches understanding of the LCA of a non-ICT product (e.g., paper) which is influenced by the availability of ICT services. Research in the second-order-impact branch of research seeks to optimize design, production, use and disposal of non-ICT products, or modify demand of non-ICT products through encouragement of substitution (decreasing consumption) or induction (increasing consumption). Green IS's first- and second-order-impact branches re-converge onto the development of EMISs; the third-order-impact branch includes research that seeks to develop goods and services with the support of EMISs.

Like Green IT/ICT, Sustainable HCI has "in" and "through" divisions: "Sustainability in Design" (first-order-impact research) and "Sustainability through Design" (third-order-impact research). Sustainability in design seeks to increase useful life of products, to facilitate transfer of ownership and to improve material recovery on disposal. The objective is to design for sustainable interaction throughout the product's lifecycle. Sustainability through design consists of the design of technology and interactive systems to support sustainable lifestyles and promote sustainable behaviour. Sustainability through design thereby supports ambient awareness; develops persuasive technology; facilitates pervasive sensing; enlists people's support in participatory sensing, and engages in formative user studies.

The framework would be incomplete without structural provision of attributes for the nodes. To date, this has been carried out for the nodes at the level denoted "Community Classification, Level 1", where "Key Differentiators in Perspective" and "Hints of Isolation of Communities" have been found useful to the declared intention of providing a novice with a map to the field. It is expected that more attributes can be extracted, at this level as well as at level 2 and others that might ensue.

2.2 Converging onto the metro area

This section shows how the process of surveying literature estimating use-phase energy consumption, led to limitation of scope to the metro area of network deployments.

2.2.1 A significant problem …

Ishii et al. [40] indicate that in the future (up to 2030), should the current architecture (as described in [40]) prevail:

1. Total energy consumption (energy consumption) of the Internet in Japan due to fixed broadband access increases at a supra-linear rate, from around 7.5 TWh in 2015 to around 22 TWh in 2030.
2. This growth in energy consumption was to come from the metro-core and core networks, collectively referred to as the backbone. In [40]'s architecture, these segments of the network contain layer 3 activities and associated optical transport.

No direct explanation is given of the cause of the projected 30% compound annual growth rate (CAGR) of average, per-subscriber, download traffic. This prompts the next two steps: corroboration of the claim of growth and in such a case, the cause of the growth. A hint of the cause may be found in a recent report by the independent firm D+R international [147].

2.2.2 … with significant solutions

D+R International works with the US government and the private sector on programs related to energy use. One program is the Voluntary Agreement for Ongoing Improvement to the Energy Efficiency of Set-Top Boxes (STBs). The program is steered by the Internet and Television Association (NCTA) and the Consumer Technology Association. It is administered and audited by D+R International. The Agreement's signatories bind themselves to meeting Energy Star standards within specific dates for a specific percentage of STBs procured by (video) service providers.

In the latest annual report on the implementation of the Agreement [41], some saliencies are striking. With respect to STB energy use of 32TWh in 2012 (predating the agreement),

- energy use in 2016 was 24.5 TWh, representing a saving of
  - 5.2 Mt CO2e emissions and
  - USD 941 million;
- total savings, counting savings from 2013 to 2016 are

- 11.8 Mt CO2e emissions and
- USD 2.1 billion.

The significance is startling: this four-year energy saving is equivalent to the annual energy use of the cities of Washington, D.C. and Chicago combined [41]. More detailed investigation is warranted of whether traffic is growing and any underlying cause. The following three sub-sub-sections are the results of the investigation.

2.2.3 An investigation of trends in traffic volume

Corroboration (1): Growth in aggregate subscriber traffic

Cisco estimates IP traffic exchanged ***over the access network*** by both businesses and consumers with:

- endpoints over managed networks and
- endpoints over unmanaged networks ("Internet traffic").

Both the mobile access network and the fixed access network are considered. Cisco considers IP traffic over managed networks to be characterized by passage through a single service provider. Without explicitly referring to quality of service (QoS), the implication is clearly that the traffic is controlled to meet the QoS demanded by the service level agreement (SLA). In contrast, "Internet traffic" crosses provider domains; typically, this traffic is delivered on the basis of providers' best effort. These two kinds of traffic complement one another and collectively are referred to as total global IP traffic. **Error! Reference source not found.**shows the development of the projections in four consecutive years.

Table I CAGR[5] REPORTED IN CISCO'S VNI OVER FOUR CONSECUTIVE YEARS

Period	Fixed Internet Traffic	Managed IP Traffic	Mobile data
2014-2019 [42]	23	13	57
2015-2020 [43]	21	11	53
2016-2021 [44]	23	13	46
2017-2022 [45]	26	11	46

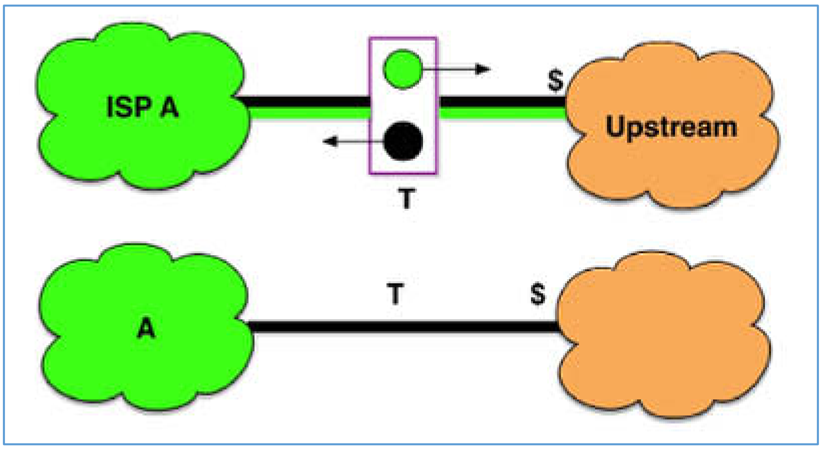
Corroboration (2): Growth in Transit Traffic

The notion of Transit Traffic is illustrated in **Error! Reference source not found.** [19]. Traffic rate is measured at regular intervals (5-minute intervals are typical). Although an ISP is shown as the

---

[5] Compound annual growth rate. Note that the figures refer to CAGR; they do not refer to the percentage share of total traffic. The key observation lies in the realization that, year after year, significant (heavy, in the case of mobile data) growth is persistently predicted.

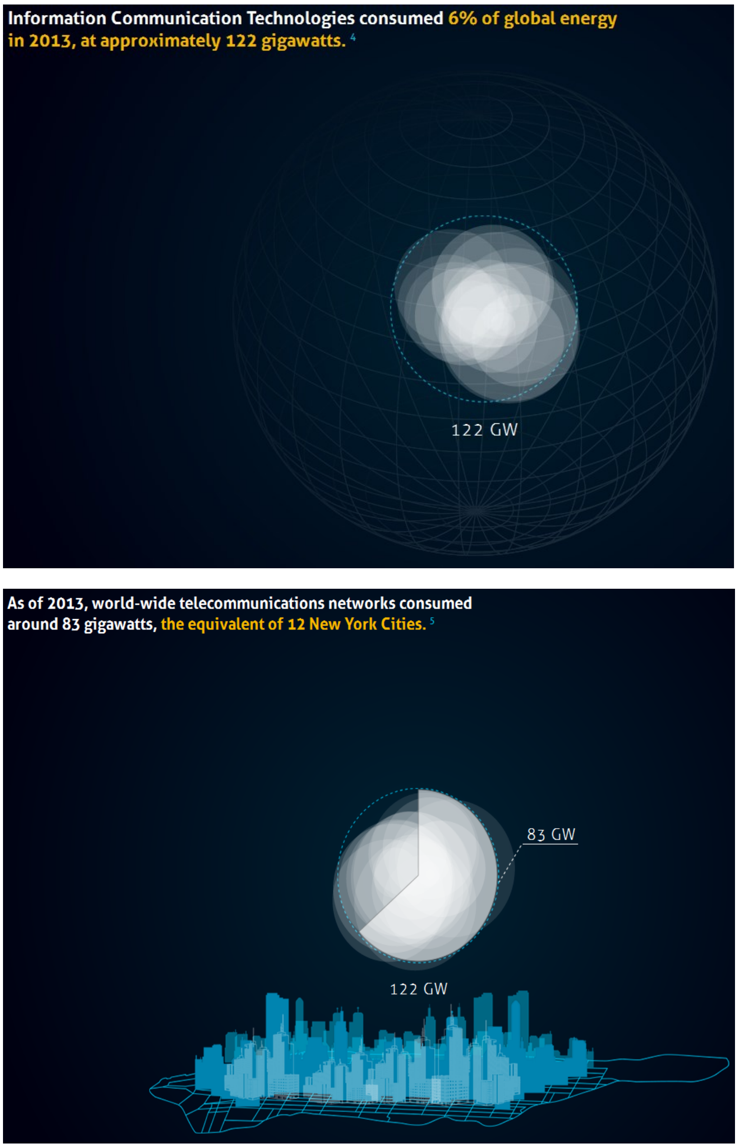
customer (the $ symbol refers to who is getting paid), the use of traffic rate metering applies widely outside the residential market. For example, co-located customers in a data centre may be charged for Internet traffic rate using some statistical measure/ The 95th percentile over a month is common. In [42], Internet traffic is quoted as increasing at a rate that typically exceeds 50% per annum; in [19], the estimate is similar, at 40 to 50% per annum. Since inter-datacentre traffic appears to be excluded from [46], this growth figure firstly serves as further corroboration of growth in demand and secondly seems to be a superset of the scope of [46].



2.2.4 Investigating energy consumption trends across network segments

A comparison of the results of a number of studies [14], [40], [47] that estimate the growth of energy consumption in the Internet, shows significant differences in the estimates attributed to the segments of the Internet. However, there is consensus on an increase in the total energy consumption. One important cause of the differences between studies lies in the difference between the boundaries applied by the respective studies.

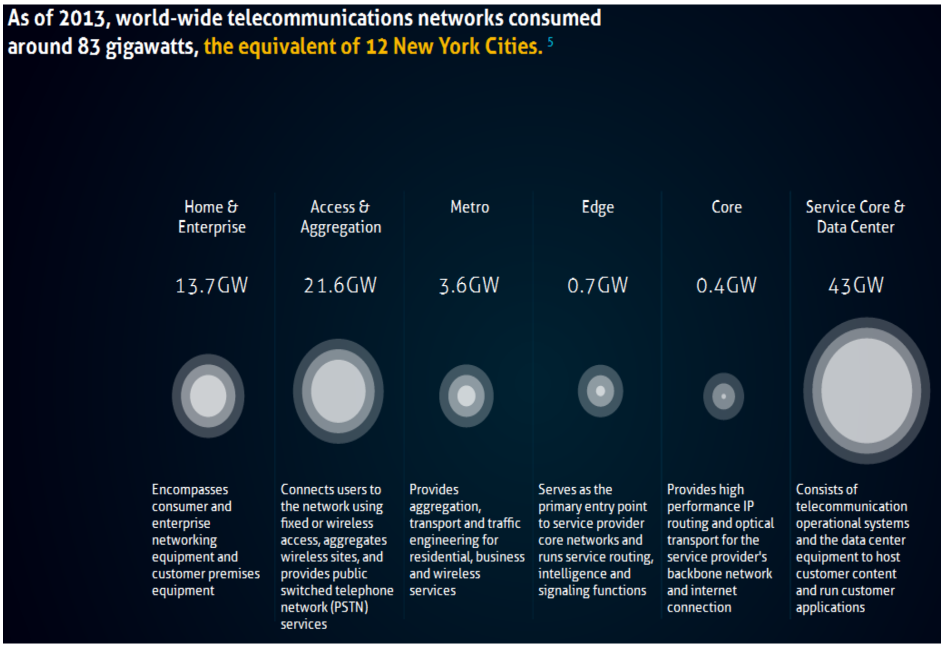
The GreenTouch Foundation has issued a white paper summarizing its findings [48]. Figs. 2(a), (b) give a first impression of the insights published by Bell Labs (now Nokia), one of the Foundation's members. Their Green Meter study targeted communications networks, which were foreseen to grow their consumption of energy as a consequence of the exponential growth in Internet traffic (see Fig. 2(a) [14]; also see Fig. 2(b) [49]).



Information Communication Technologies consumed 6% of global energy in 2013, at approximately 122 gigawatts.[4]

As of 2013, world-wide telecommunications networks consumed around 83 gigawatts, the equivalent of 12 New York Cities.[5]

Fig. 1. (a) First & second of four graphics [50], [51], [52], [53] showing a snapshot of networks' share of the impact



2.2.5 A compilation of sources, identifying the cause of growth in Internet traffic

Cisco: Dominance of Consumer IP Traffic by Video Traffic

In [42], Cisco's VNI distinguishes between managed IP and Internet (unmanaged) traffic. Managed IP traffic is private to a single service provider's infrastructure. For example, managed video IP traffic is content from the service provider's catalogue, which is provided to its subscribers and is distributed from a head end or central office towards the consumers at the termination of the access network. Cisco's VNI states that, if only unmanaged (Internet) consumer traffic is taken into account, consumer Internet video traffic was 64% of all consumer Internet traffic in 2014 and projects growth to 80% in 2019. If both managed and unmanaged (global) consumer IP traffic are taken into account and all types of video considered (Internet video, P2P file sharing and managed IP forms like IP transport of TV and VoD), video was projected to take up 80 to 90 percent of the global total by 2019.

Sandvine: Dominance of fixed access networks by real-time entertainment

From the perspective of subscribers of fixed access connection, Sandvine [54, p. 5], [55, p. 2] states that real-time entertainment occupies 67.40% (2013) and 70.40% (2015) of downstream bytes respectively in North America, exceeding the second largest component (web-browsing, 7.01%) by an order of magnitude in the most recent report. For wireless access, in [54, p. 9], [55, p. 6], downstream bytes occupy 39.91% (2013) and 40.89% (2015). Real-time entertainment is again dominant, with social networking coming in second at 20.53% and 22.06% respectively. The dominance of real-time entertainment in fixed access networking is explained in [42]. In 2019, Wi-Fi traffic is expected to comprise a lower percentage of managed IP traffic than Internet IP traffic, due to the large portion of managed IP traffic attributable to an IPTV provider's CPE. Without any further evidence, a claim that this is the same reason behind the dominance of real-time entertainment in wired devices' downstream consumption cannot be made; there are steps missing in the logic. It does however provide a hypothesis which is likely to be resolved with further investigation into this and other sources. At this point in the development, however, these investigations are not warranted as the evidence of the dominance of real-time entertainment is sound. Indeed, a later edition of Sandvin'’s Global Internet Phenomena report [56] dedicates its executive summary exclusively to video traffic; the report shows video as consuming 60% of downstream traffic – a further 2% increase over 2018.

William B. Norton: Evolution of the Internet towards localized traffic growth to accommodate video

In [19], a series of evolutionary steps are described and are used to explain the change in the model that graphically describes connectivity in the Internet. This model can be used to explain the Netflix-Comcast and Netflix-Verizon conflicts. The conflicts are visibly rooted in peering arrangements that Netflix formed with intermediary ISPs. Netflix has expected the access network providers to peer (freely) with its own CDN at the metro-core segment in the same manner as it has with intermediaries like Cogent [57]. Netflix would like to see its role evolve in the manner depicted in Fig. 3. The illustrations are from [19], but the indication of the change sought by Netflix in [57] is my own application of the model in [19]. Content providers are motivated to (a) keep away from backbone transit providers and (b) reduce the number of hops between their content and the access networks. Netflix has been to date unable to reach this objective with Comcast and Verizon and is paying these ISPs for peering at the metro-core segment, in order to fulfil its customers’ expectations of flawless video. It has been successful with cable broadband providers Cablevision in the north-eastern US and Grande Communications in Texas [58]. Regardless of the type of peering adopted, this peering (free or paid) has the effect of loading the metro-core segment with traffic that would otherwise have been routed through other parts of the Internet. Two techniques used by Netflix in the deployment of their content delivery network (CDN) have been (a) peering at Internet Exchange Points (IXPs) and (b) installation of their Open Connect Appliance (the CDN server at the metro-core) inside Internet Service Providers’ (ISPs’) datacentres [59].

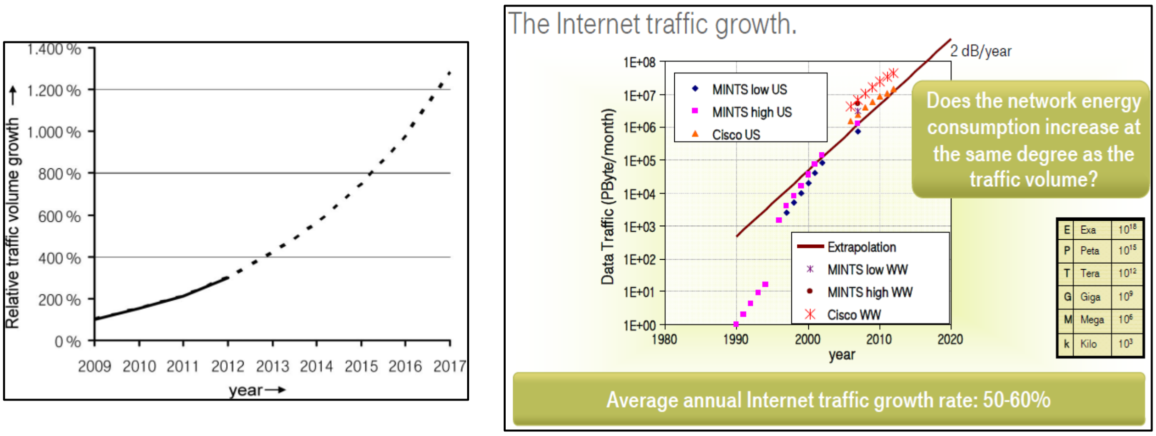


Fig. 2. (a),(b): Internet Traffic Growth [14], [49]

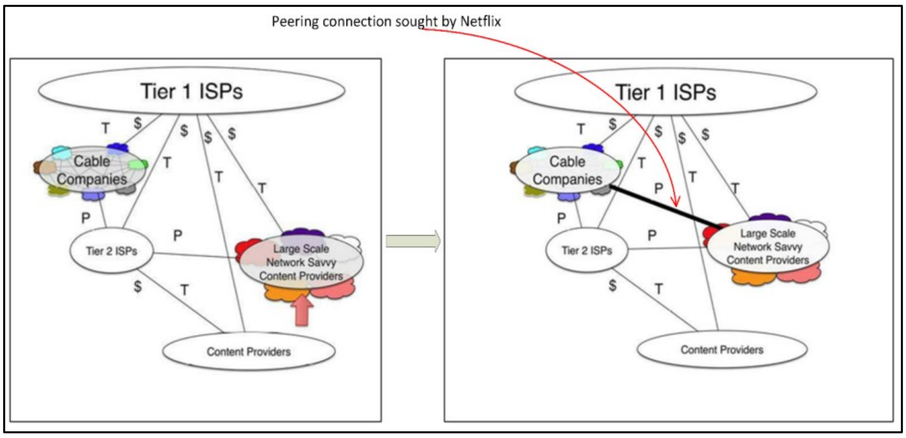


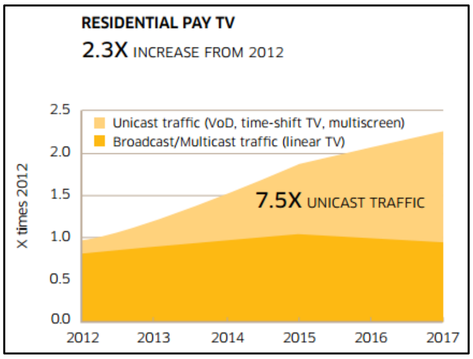
Fig. 3. Content Providers want to keep their traffic off the Tier 1 backbone networks and close to access networks. T stands for paid transit and P stands for free peering. Paid peering (PP) is not shown in this model [19].

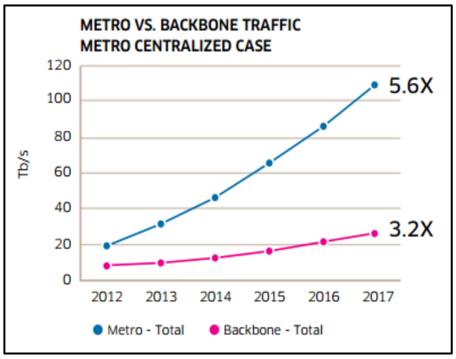
Alcatel-Lucent Bell Labs: Corroboration of metro-core video traffic growth

In [46], video traffic in the metro-core in 2017 was predicted to grow to 7.2 times the traffic in 2012. Video traffic is described using the same distinction made in [42] between managed and Internet traffic. Managed traffic is referred to as "Pay TV" and is divided into traditional, linear viewing to a set-top box (STB) and non-traditional viewing like video-on-demand (VoD), network personal video recorder (PVR) and time-shifted TV on IP-capable devices. Internet video comprises the complement to Pay TV. **Error! Reference source not found.** [97] shows the projected growth in managed traffic.

Alcatel-Lucent Bell Labs: Corroboration of localized traffic growth

In [46], total traffic in the metro-core in 2017 was predicted to grow to 5.6 times the traffic in 2012. Apart from video, datacentre interconnection traffic is projected to grow rapidly, reaching, in 2017, 4.4 times the traffic of 2012. 75% of this traffic is expected to remain intra-metro; the remaining 25% is expected to cross over into the core. **Error! Reference source not found.**[46] compares the p rojected growths by segment.





2.3 The primary driver of research: growth of IP video traffic energy consumption

The growth of energy consumption by the network to support video traffic is recognised by researchers, who have been tackling the problem [46], [60], [61], [62], [63], [64], [65], [66], [67], [68],

[69], [70], [71]. From these works, development of cache architectures emerges as an important approach to controlling energy consumption but others exist and yet more are emerging within the general thrust towards "future networks" (ITU-T Y.3001 [72]). Radical and reformist approaches are distinguishable: radical approaches employ dynamic and reactive control while reformist approaches seek to improve the caching of content and are characterized by investigation of the reduction of the length of the path between source and destination of IP traffic.

2.3.1 Reformist approaches

Reduction of source-destination path length is premised upon the insertion of caches closer to the point of consumption than the Internet core (where CDN points of presence may be found). Jayasundara et al. [69] succinctly represented the opportunities for improvement in an early work that identified the various insertion points of caches in the service delivery chain. A first impression might be that the problem space of cache placement is diminishing since the root cause of restrictions in cache placement – cost and speed of storage – is fading in severity. Further reflection reveals that the problem space is changing as technology evolves, rather than diminishing.

- The proliferation of caches leads to increased energy consumption. Whereas cache placement effectiveness used to be balanced with cost and speed of storage, it must now be balanced with the cache and cache-support infrastructure's energy consumption.
- Caches vary in size. While a Netflix Open Connect Appliance (OCA) running NGINX CDN software and multi-terabyte storage may well sit in a datacentre occupying rack space, the cache in a set-top box (STB) is unlikely to be more spacious than one SSD's worth of data. A high-level description of the problem matches that which has been addressed in a computer system. The fastest cache (in the STB) is smaller than the next fastest cache deployed in the access segment (e.g., the DSLAM, the OLT, the eNodeB and the gNodeB); this in turn is smaller than an OCA in the metro-core segment.

Some examples of the reformist category of approaches are:

1. OTT-ISP collaboration [67]
2. Exploitation of fixed – mobile convergence (FMC) [68]
3. "Metro-server" caches (edge compute caches) [46], [70]

2.3.2 A radical approach through Software-Defined Networking.

In the guest editors' introduction [73] to an issue of Computer Magazine focusing on Software-Defined Networking (SDN), this domain was referred to as "the emerging second wave" of Cloud Computing (the first wave was identified as "server centralization and virtualization"). The excitement about SDN may perhaps be most representatively summarised in the reflection that its development is guided by the growing need to create abstractions of the network that facilitate the development of software to manage the network.

The problematic interweaving of network control with network traffic

Bolla et al. [74] identify an "ossified TCP/IP structure" as a factor that has restricted the evolution of the Internet. Corroboration and more insight are found in Stallings' [75] and Shenker's [76] analyses. Traditional network control is implemented through a complex mix of protocols that are focused on aspects such as routing, QoS, security and mobility. The difficulty is that the Internet Protocol architecture provides abstractions for data delivery but not for network control. At layer three, IP architecture provides addresses that identify source and destination and a maximum time-to-live. At layer four, it provides sockets for multiplexing and de-multiplexing onto a single address, as well as the means to synchronise the sender and the receiver with regard to the data stream between them. At the (OSI-)application layer, protocols define means for user applications (not network control applications) to exchange data over the Internet. The substance of the difficulty lies in the ingraining of a data-plane-abstraction architecture into applications for network control. This ingraining has constrained the evolution of networking to conform with the layers of the Internet Protocol architecture. It has coupled applications for network control into data delivery paths. The network application typically receives control data in packets that are exceptions to the general case of packets destined for forwarding through the packet switch. The network application exercises control through modification of data plane structures within the switch.

The problematic condition of static network management

In its current form, dynamic control of networking devices is not scalable as the effort required and subjection to error grow at a fast rate with increase in the number of devices to manage. Manual procedures, vendor-specific methods of configuration and the incompatibility amongst network equipment vendors' scripting languages are some of the difficulties that limit scalability. Solutions exist that address the difficulties of the configuration phase, to some extent. For example, NETCONF (Network Configuration Protocol, RFC 6241) is an intermediate layer that serves to interface a network management system (NMS) with network devices at the communication layers of transport, message format, device operations and content. Nonetheless, it has been noted [77, p. 118] that even these solutions afford interfaces that are unable to offer a comprehensively dynamic interaction with applications that require real-time bi-directional control with the network elements (e.g., packet switches and firewalls).

Decoupling application architecture from TCP/IP architecture

Networking can be made more agile. "Agility" is a term that is used to describe the ease with which a network may be modified to fulfil the requirements of the services which it carries. The prevalence of OpenFlow has heightened awareness that the configuration sub-phase of implementation can be shortened. For example, for the networking devices (e.g., packet switches) already in the field of operations, configuration could be automated if the physical resources were to be abstracted into logical resources accessible through a programmable interface. OpenFlow is one such means that provides this abstraction. NETCONF, supported by YANG ("Yet Another Next Generation" – RFC 6020) data models, is another.

A more concerted solution to improving network agility would transcend individual device configuration and encompass as many phases of the network's lifecycle as possible. The Open Data Center Alliance proposes "Model Management", which is described [78] as the implementation of conceptual changes to a network by modifying its model – rather than through the reconfiguration of "individual network elements". For example, an arbitrary network topology may be modelled as a crossbar (Fig. 4 [76]). In terms of layer 3 virtual private networks (L3VPNs), this is indeed the model adopted by communications service providers (CSPs). Customers connect to the network using a link between their own layer 3 (L3) routers (the customer edge – CE – devices) and the CSP's own L3 routers (the provider edge – PE – devices). The connection to the crossbar is the link between the CE router and the PE router. The customer is only concerned with obtaining a route, at every CE, from the peering PE, towards the desired cEs. The role of network agility is automation of the task of setting up a path through the CSP's (abstracted) network nodes, that meets the service level objectives (SLOs) specified in the service level agreement (SLA).

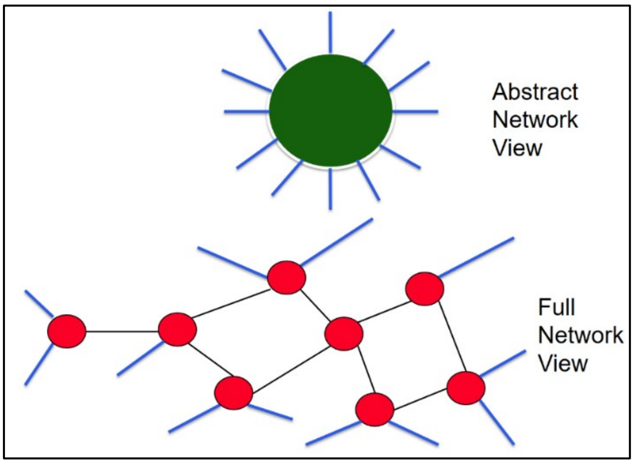


Fig. 4. Abstract views simplify & are relevant to a specific network application(e.g. access control, end-to-end connectivity) [76]

Beyond the specific case of automating the creation of L3VPNs, this model is applicable to at least all other contexts that demand (layer X – whatever layer, that is) end-to-end connectivity (an N-port virtual switch) and, of course, access control. A network application is used to configure this model. Coherent configurations are then pushed out to all devices involved in the application's deployment. In this vision, the networking is software-defined:

- a management-plane component is used to configure functional requirements through a network model;
- this model is transformed through (inversion of) a number of abstractions into a set of device configurations that collectively meet the functional requirements of the application;
- the data plane switches are programmed according to the prescribed configurations;
- modifications to the configurations are effected during the network's operation to reflect real-time conditions.

Decoupling network functions from hardware

Networks are evolving into flexible and programmable "softwarized", virtualized infrastructures. Deployment of network functions is undergoing the paradigm shift from physical network functions (PNFs) to Virtualized Network Functions (VNFs); this has come to be known as Network Function Virtualization (NFV). Such (softwarized, virtualized) infrastructures confer ***agility*** to systems they constitute: they comprise the dynamic, reactive data plane employed in SDN. At least two thrusts can be detected in any discussion on the significance of the SDN + NFV paradigm.

1. Strongly integrated paradigms that explicitly refer to virtualized infrastructure, like 5G [79], [80], promote their adoption.
2. The extent and relevance of the control plane is broadened by standardized architectures that equip the control plane with uniform interfaces that exploit extant and emerging (including ***green***) capabilities.

NFV involves moving into a hardware and software ecosystem that exploits ***general-purpose computer systems*** (or "COTS" – commercial-off-the-shelf- hardware and software) to the greatest extent permissible: this is a declared objective[81, p. 8]. In this ecosystem, the general-purpose computer system is a host for the network functions, replacing the specialized hardware that characterized physical network function deployments. Moreover, the general-purpose computer system is a bundle of resources, by virtue of the facilities conferred by machine virtualization and containerization. The primary benefit here is that the unit of computing is no longer inextricably bound to a physical platform, but can migrate with (nowadays) little effort from one physical platform to another. Henceforth, where used, the term virtualization platform[6] refers to the hardware and system software (for machine virtualization and/or containerization) upon which the VNFs are instantiated.

---

[6] Within NFV MANO, the virtualization platform is referred to generically as "Network Functions Virtualization Infrastructure" (NFVI), and a single physical instance of NFVI is referred to as an NFVI-Node.

NFV adds a new dimension to the problem of power use: ***the impact of virtualization technologies on power consumption in public telecommunication networks (PTNs) is still unclear***.

- On the one hand, there is a general belief that Network Functions Virtualization (NFV) should result in reduced energy consumption, owing to consolidation of resources and increased flexibility in turning unused hardware (HW) on and off as needed.
- On the other hand, it is also true that "the massive introduction of general-purpose HW enabled by NFV would tend to increase power requests with respect to specialized HW solutions" [79].

Therefore, there is a need to operate power-aware management and control mechanisms in these environments. At the same time, it is necessary to limit the complexity of these mechanisms and the level of human intervention therein, to keep Operational Expenditures (OPEX) within reasonable limits. One approach to understanding the impact consists of comparative analyses of implementations of infrastructure, with and without virtualization. This approach is taken in [82], where the evolved packet core (EPC) is studied. This work shows that the virtualized implementation is indeed less energy efficient. Unfortunately, the scope of virtualization and containerization within the converged wireless and wireline infrastructure is very broad and consideration of a single "use-case" [83] cannot be generalized to an overall statement. A network-operations context is required: I suggest that it is what has come to be known as the "telco cloud".

"Telco cloud" is an evolving notion that evokes a number of common terms in attempts to describe it. Virtualization, software-defined networking (SDN), automation and orchestration are four such terms. Other prominent terms are edge computing, containerization, microservices and resilient infrastructure [84]. I suggest three key observations that organizes these terms into a coherent image of the telco cloud.

1. The telco cloud is, fundamentally, a hybrid cloud:
    a. self-sourced virtualization and containerization, and
    b. out-sourced (public cloud) containerization.

   The complementary collaboration of the CSP's network, compute and storage infrastructure, with that of global providers of infrastructure and applications, is manifested well in [85]. A distributed cloud infrastructure operates at network (transport and interconnect) junctions. It includes (cloud) infrastructure owned and operated by the CSPs, by public cloud providers, and by enterprises which consume their joint service.
2. The telco cloud serves both internal and external clients [84].
    a. Internal use can suggestively be referred to as the *IT Cloud* [86]. This consists of applications specific to CSPs: operational support systems (OSS) and

business support systems (BSS), as well as more general applications, like customer relationship management (CRM).
   b. External uses are growing organically on the basis of use cases seeded by ETSI [83] and the 5GPPP [87].
3. The service-based architecture (SBA) of the 5G Core is a good fit with cloud-native computing. Containerization is distinctively central to cloud-native computing [88], [89]. The Cloud Native Computing Foundation explicitly identifies containers as components of the approach to the concept of cloud native computing [90]. There is a clear drive towards use of containers in lieu of virtual machines as the operating environment for network functions [88], and the 5G Cor''s SBA provides a clear scope for employing containers.

The real estate where the VNFs may be deployed have been labelled further with the term "network functions virtualization point of presence" (NFVI-PoP). These may be:

1. **datacentres**: here, the real estate referred to consists of points of presence (PoPs) such as metro-core PoPs at the near edge;
2. **within softwarized and virtualized network**s: here, points of presence such as central offices (cOs) and sites even deeper into the edge such as remote radio head (RRH) sites and roadside cabinets, are indicated.

2.4 A theoretical framework: Attributional Life Cycle Assessment

Attributional life cycle assessment (henceforth referred to as LCA) provides an overarching theoretical framework within which to interpret this work. In this sub-section, the grounds of this study are stated and the principal objects of the study are aligned with the major abstractions of LCA [91], [92]: ***product, system boundary, function*** and ***functional unit***.

2.4.1 The grounds of this study

This study is grounded in the need (expounded in sub-sections 1.1 and 1.2) to describe the metro area network in a manner conducive to universal accord. Within LCA, this supports the pursuit of ***parameterized modelling***, which was recognized early in LCA development as a "need … to reflect technical scenarios and parameter variations which would provide a sense of the importance and influence of the possible consequences of future development and technology or supply chain changes" [93].

Parameterized modelling is obtained through detailed understanding of ***unit processes*** throughout the phases of the LCA. Notably, the first phase of the LCA (goal and scope definition) demands (in scope definition) an explicit ***system boundary***. The system boundary takes on particularly onerous ramifications where the goal of the LCA is that of comparative assertion ("environmental claim regarding the superiority or equivalence of one product versus a competing product that performs the same function" [91]). That is: unless comparisons are fair (notably, through transparency), then they defeat their goal and turn into obfuscations. I propose to support fairness in goal definition through facilitation of transparency ("open, comprehensive and understandable presentation of information" [91]), in turn through adoption of readily available and recognized ***reference points***, where available, and identification of ***potential points for standardization*** where otherwise unavailable.

2.4.2 Interpretation of the major abstractions of LCA

Following Schien et al. [94] (where the product was online digital news), I identify the ***product*** as a digital service, delivered over a telecommunications network. Schien's analysis is limited to the ***use phase*** of the digital service's (i.e., the product's) life cycle. This limitation requires a concomitant limitation of the concept of ***system boundary***, as the original definition [91] introduces it in relation to a ***product system***, and the latter defines the scope of the life cycle (not part thereof) of a product (again, see [91]). In other words: the product system can be understood as an encapsulation that serves as the analytical reference of the product's life cycle.

System boundary

In the LCA framework, the system boundary demarcates the product system. Furthermore, it gives meaning to the directions "in" and "out" used with an LCA's flows: these directions are only useful in relation to some enclosed space – or scope. The limitation on boundary is endorsed by Schien's graphical instantiation of ISO 14040's system boundary. The graphic indicates energy flows inward, but excludes embodied energy; the latter is used during ***manufacture*** of the equipment that carries the data flow during the use phase.

It is reasonably arguable that this limitation on boundary is grounded. A first justification may be found in [95], where the system boundary is shown to span a range of possibilities:

- at one end, the system boundary includes production, use and disposal of all of an organization's products;
- at the other end, the boundaries include a single production facility (hence, the production phase only).

Moreover, in [93], the inclusion of several life cycle stages is recommended, not mandated (let alone the entire product's life cycle). Limitation to a single phase can therefore be tolerated. Schien's system boundary is shown in Fig. 5 [94, Fig. 1].

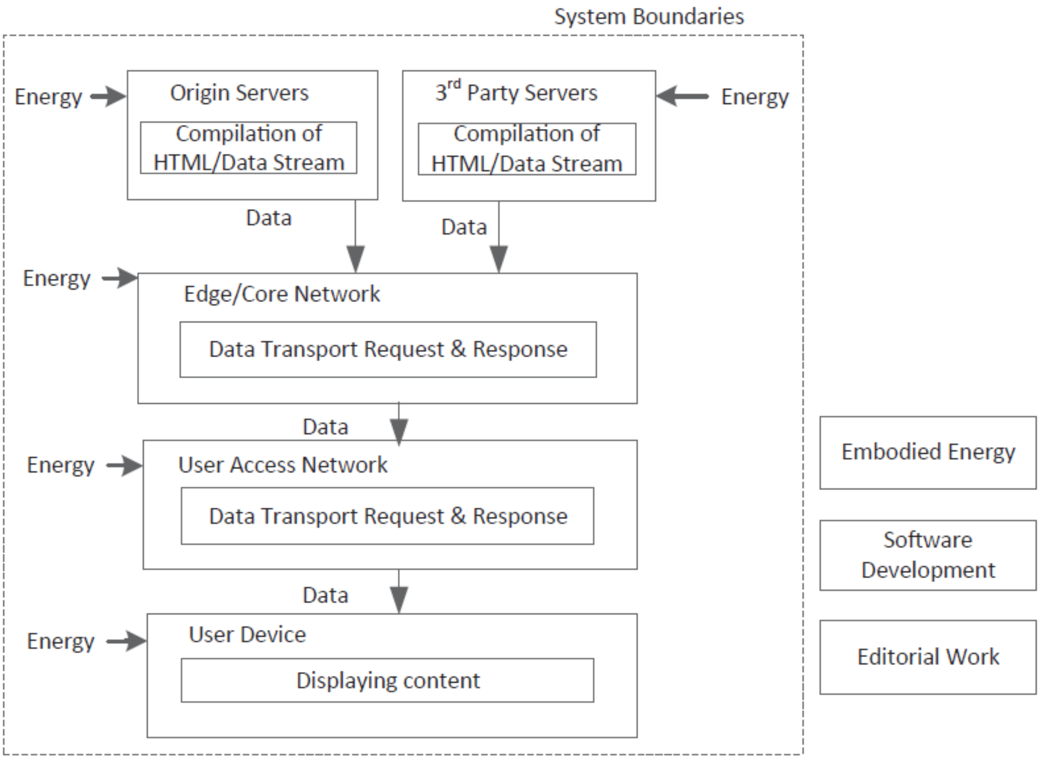


Fig. 5. System boundary around the use phase of a digital service

Unit process

I now turn to an interpretation of another of the LCA's major abstractions: the unit process(es). Now: the product system abstraction elicits the unity of a potentially highly disgregated and disparate set of elements that produce, transport, use and eventually dispose of a product. ISO 14040 appropriately uses the phrase ***collection of unit processes***:

- "unit processes" is appropriate because it emphasizes the self-sufficiency of a unit and the flow-like behaviour of a process;
- "collection" is appropriate because of the inherent looseness in the term "collection": the relationship between processes is the flow of (intermediate) products to successors in the life cycle.

While Schien does not explicitly identify unit processes in [94], comparison with the general product system [91] gives some clear indications. The unit processes are subsumed, without explicit identification, within the major segments of the chain of hardware and software between, at one end: the dynamic compilation of the HTML – (hypertext markup language) and data – stream, to, at the other end: the display in the user device. The major segments shown are few enough to list readably: origin and 3rd party servers, followed by edge/core network, followed by user access network, followed by user device. Both the general model and Schien's model directly refer to the energy flow into systems of unit processes. Schien's statement of the scope of the segments is problematic: it suggests a universality of interpretation, but I deal amply with the variety even of such coarse modelling in Chapter

4. Disambiguation efforts there (i.e., in Chapter 4) identify ***four*** such coarse models, organized in ascending order of granularity. Moreover, with generalization in mind, parameterization was limited to (standardized) reference points and had not identified the processes within the segments. In Chapter 5, this limitation is partially addressed through technological instantiation. Although obsolescence is baked into such instantiations, they are valuable as they provide baselines upon which to develop updated models. Through such means, unit processes can be defined (or, at least, described) with far less uncertainty in the ensuant product system (or part thereof).

Function and functional unit

The last of the major abstractions of concern are the function and functional unit. Recently, Shi et al. [96] have drawn attention to the (non-compliant) selection, in smartphone LCAs, of the smartphone itself as the functional unit. Shi points out a lack of resolution into “the impact of product functionalities”, and “variation in impacts depending on how users use different functions”, and calls for quantifiable function of the smartphone as functional unit. On the other hand, Schien defines the functional unit as ten minutes of time browsing multimedia content (text, images, audio and video). This is, indeed, a quantifiable function of the product, as follows. The product is the digital service and its function is to inform viewers through (multimedia) presentation of news content. The functional unit is meaningful because ten minutes of browsing represent a session of human interaction. This understanding of function links well with, and expands, two observations made earlier:

- the telecommunications network enables a product mix;
- good characterization of digital service use phases facilitates a more accurate characterization of the telecommunications network’s use phase.

With these in mind, the function of the telecommunications network may be thought of as ***a complex product with multiple functions***: the functions of the digital services which it enables. The mathematical foundations of this approach have been established in [97], [98]. The function matrix of the telecommunications network can be written as follows:

$$F = \begin{bmatrix} | & | & \cdots & | \\ f_1 & f_2 & \cdots & f_n \\ | & | & \cdots & | \end{bmatrix} \tag{1}$$

where the column vectors f_j , $j \in \{1,2,\ldots,n\}$ each represent a digital service. The element in row i , $i \in \{1,2,\ldots,S\}$ denotes energy input to a specific unit process out of the set of ***S*** unit processes that are included within the system boundary. Therefore, the unit processes are segments of the path, from source to destination, through the telecommunication network under study – in this case, the metro area network. This completes the alignment which I set out to achieve. The core impetus henceforth will be to define this path to facilitate future LCAs of digital services and metro area networks.

2.4.3 Extant concretizations of the major abstractions

In the previous sub-section, an important step was taken towards interpretation of the telecommunications network from the lens of the LCA practitioner. The unit process, product system and product were instantiated, albeit at a level that requires further concretization before CSPs and their technology vendors may be inseminated with the grasp necessary to productively and generatively discuss implementations of Life Cycle Assessment. Unless CSPs and vendors "get the feel" for what LCA stakeholders are preaching, it is unlikely that further inroads into greener digital services will be made by joint LCA – CSP task forces. The disciplinary divide between the LCA practitioner and that of Information and Communications Technology (notably those spaces of ICT where formation of the CSP's network engineer is cultivated) reflects two ontological planes that require transformative mappings to link concepts, categories and the ensuant properties on the respective planes. **This sub-section addresses this need for concretization.**

Transformation onto the ontological plane of the ITU-T's Next-Generation-Network

Rigorous interpretation of Schien's boundary requires ***location*** of the digital service within the physical context of the use phase of its life cycle. Intuitively, this process of location supports both the interpretation of the service's use phase as well as that of the telecommunication network. A broad spectrum of products (digital services) is enabled by a telecommunications network; indeed, a telecommunications network may be said to enable a product mix, the QoS of which varies widely [99], and may range:

1. from inelastic, high bandwidth, high-availability, low-latency, symmetrical (equal upstream and downstream requirements) applications like video conferencing,
2. to elastic, low bandwidth, outage and latency tolerant, asymmetrical applications like managed downloading.

By categorizing traffic according to digital service, its diversity is sorted into service silos that support the ***attributional*** approach to the network's LCA. Therefore, a well-characterized use phase for the digital service facilitates a more accurate characterization of the telecommunications network's use phase. This indirect approach to the network's use phase is highly amenable to the network's life cycle assessment.

Th47odelliss of ***location*** of the service within the telecommunications network (and any other supporting infrastructure) is evidently facilitated by the availability of detailed mapping of the network. Moreover, ***standardized*** mapping directly addresses the challenge of implementational diversity and directly contributes towards achieving reconciliation of architectural paradigms. Obtaining a recognizable system boundary is thus scaffolded by the use of standardized representations of the telecommunications network. A thorough investigation of SDOs' works is warranted. This must extend beyond compilation of their independent mappings, into cross-comparison of any such mappings, to establish the breadth and depth of their adoption.

Until now, appeal has been made to the intuitive sense of the term "segment", to permit deferral of a more formal explanation until such time as the weight of its relevance becomes apparent. By now, this should be so, and the relationship between the segment and the system boundary can be explored in more depth. Fundamentally, and in essence: ***a segment is an entity of organizational structure***. This is the core of the "semi-formal definition" of the term, made in the ITU-T's Recommendation Y.110 [100]. Y.110 appears to be the primary source upon which are based all subsequent references made by the ITU-T (see, for example, how deference is yielded to Y.110 in ITU-T Y.140 [101, Sec. 2]). The definition is used in the context of a variety of modelling schemes, and in each scheme the commonality ("is common to") is emphasised. The definition given is that:

> *"[a] segment is part of one role,*
>
> *owned and operated by one player,*
>
> *part of one (and only one) service provisioning platform, and*
>
> *part of one domain, and*
>
> *is composed of a well-defined set of functions."* [100, Para. 6.1.2.9, 7.1.2.9 and 8.1.2.7]

Segments are bounded by ***reference points*** (RPs); this follows from the assertion carried in Y.110 that an RP lies at the interface between two functions (both functions' sides of the interface are specified). Specific reference is made to "transport telecommunications reference points"; these "can transparently support other logical interfaces including application protocols, middleware protocols, and even the control protocol between the base functions and the network control functions". This draws the following parallels with LCA terminology:

1. **Comparable hierarchy in ontologies:** Just as the product system lies above the unit process in the hierarchy of objects in LCA's ontology, so is the segment above the function in the ontology of telecommunications.
2. **System boundary – reference points equivalence**: Just as the product system is bounded by a system boundary, so is the segment (and therefore the domain) bounded by reference points.
3. **Unit-process – function equivalence**: Just as there are unit processes within the system boundary, so is the segment (and therefore the domain) composed of a "well-defined set of functions".

The hierarchy of objects in telecommunications networks is domain – segment – function, while that of LCA is product system – unit process. The division of the domain into segments presents an opportunity for further resolution of the product system. The understanding (stated previously) of the unit process may now be improved by using standardized terminology. The previous use of "segment" was intuitive; now, its use will follow the ITU-T's modelling. Therefore: ***the unit processes are sequenced functions, organized into segments of the path, from source to destination, through one or more domains of the telecommunication network under study – in this case, the metro area network***.

Resolving unit processes into tangible artefacts of the NGN

While establishing the primary driver of research, it was seen that approaches to controlling energy consumption may be divided broadly into two camps: a reformist one and a radical one. There, it was claimed that "radical approaches employ dynamic and reactive control", with the objective of minimizing energy consumption within the constraints dictated by the service-level objectives. The role of software-defined networking in the development of the radical approach was outlined.

Segment Routing (SR) [102] is a key enabler in the provision of this dynamic connectivity that is central to the radical way. Perhaps the two most important (if not principal) reasons for its centrality are the following.

1. It supports computation (on SDN controllers) of SR domain paths, by exportation of link state outside the scope of interior gateway protocols (IGPs), through an extension to Border Gateway Protocol – Link State (BGP-LS).
2. It avoids complexity by integrating seamlessly with extant data plane technology (notably, MPLS (multi-protocol label switching) and iPv6) and control plane technology (e.g., both IS-IS (Intermediate System to Intermediate System) and OSPF (Open Shortest Path First) support it, through TLVs (type-length-value records) and extensions).

The first (item 1, above) is really the raison ''être. Classical IGPs (e.g., IS-IS and OSPF) are good at finding shortest paths. Traffic engineering (TE) seeks alternatives: less congested paths, for example, or links that, at the ***specific*** time of investigation, have the current lowest latency, or disjoint paths through a domain. TE does not discard shortest paths; indeed, its holistic operation is predicated upon the availability of a link-state IGP. However, TE's purpose is to seek paths that have alternative objectives (as just indicated), and SR achieves this. The second reason is its key to success. CSPs seek a good return on investment (ROI), and a new technology that requires discarding an investment will be less favoured than one which can replace it organically. The first part of this second reason – "avoids complexity" – relates to the CSP's ROI on human resources. It also relates to the availability of the CSP's services. Networks that deliver services while simple to operate, are inherently better disposed to that most desirable high-nines availability. Furthermore, SR is well-poised to support Service Function Chaining (SFC) [103]. The process of instantiation of a service function chain includes a Service Function Path (SFP). The SFP traverses an ordered list of network services – or service ***functions***.

Now, Segment Routing's importance to the energy analyst lies in its architecture's intrinsic support in bridging the disciplinary separation between an LCA practitioner and a CSP's network engineer. A comparison now follows, of the relevant sub-space of the ontology of SR with its relevant counterpart in LCA, and relates the concepts and categories therein, starting from an investigation of the relationship between the service function and the same term as used in the theoretical framework (eq. (1) reproduced below).

$$F = \begin{bmatrix} | & | & \cdots & | \\ f_1 & f_2 & \cdots & f_n \\ | & | & \cdots & | \end{bmatrix} \tag{1}$$

Equation (1) shows an $S \times n$ matrix, representing the n digital services that a telecommunications network provides, through "the set of S unit processes that are included within the system boundary". It was seen that "the unit processes are segments of the path, from source to destination, through the telecommunication network under study". On the other hand, the "service functions" alluded to in the context of SFC are network functions like firewalls, load balancers, deep packet inspectors, NAT64 [104], etc. These functions may be handled linearly, cyclically, and may even involve flow multiplication (copying of packets) or division (e.g., load balancers). Now, SFC is supported through an ordered list of Service Function Forwarders (SFFs) and Service Function Instances (SFIs), encoded in the SR header (SRH, for SR-MPLS) or in the iPv6 destination address (for SRv6). A simple linear chain is illustrated in Fig. 6.

This first analysis relieves a limited similarity between LCA and SFC frameworks, which may be profiled as follows.

1. On the one hand: the LCA's product is produced in a product system consisting of a series of unit processes.
2. On the other hand: the telecommunication network's digital service is produced in a service function chain, consisting of a series of service functions (identified through an ordered list of ***segment identifiers (SIDs)***).

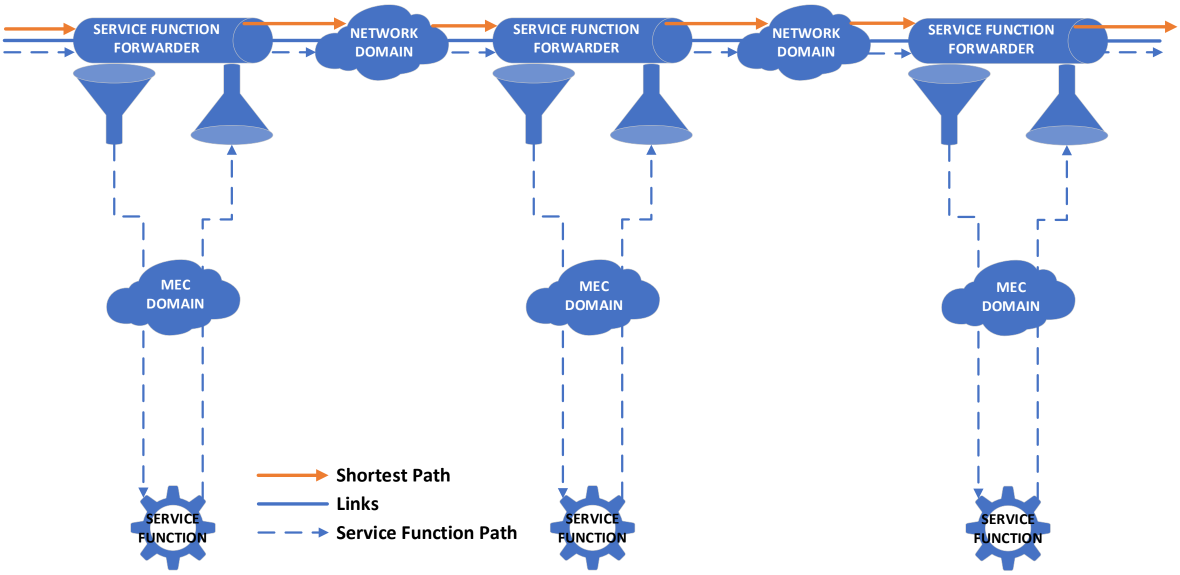


Fig. 6. Simple linear chain of service functions

The comparison relieves another similarity.

1. On the one hand: the unit processes have been identified as functions in segments of the path (this latter use of "segments" is from ITU-T ontology) through the telecommunications network from source to destination.
2. On the other hand: the service function instances have been identified through SR SIDs that compose the SFP from source to destination.

While the concept of segment in SR ontology [102] is somewhat broad, it is helpfully specific in that "segment" is bound to either a topological instance or a service instance. That is: a segment is an imperative (instruction) that drives a network element (NE) to steer a packet either through a topological instance (which may be a real link or a virtual one), or through a service instance. Simple examples include: a specific (real) adjacency of an NE, a specific tunnel which an NE has defined in its control plane and a specific device (which may be a switching NE or a service NE). The ordered list of SIDs used to express an SFP in SR, is readily comparable to the column vector in equation (1), both in the (rather obvious) mathematical sense, as well as in the physical correspondence between the underlying realities.

One part of this latter relationship between SID list and segment sequence bears further elaboration. Consider those SR segments that do not relate to the service functions (the firewalls, load balancers, etc.). These are the SR segments that relate to switching, routing and transmission. The explicit identification of these segments is not surprising, given the operational context (network operator domains (NODs), or CSP domains). However, this part of the path might be overlooked in a culture of analysis that is often inclined towards abstraction and transparency. Admittedly, reference in a SID list does not immediately solve every problem, but it does contribute to solving the problem of assurance that every part of the network is accounted for.

The similarity may be tabulated as shown in Table II .

Table II SIMILARITY BETWEEN LCA AND SFC FRAMEWORKS

Life Cycle Assessment	**Service Function Chaining and Segment Routing**
Product system	Service function chain
Product	(Digital) Service
Unit process	Service function; any other SFP component with SID
Column vectors $f_j, j \in \{1,2,\ldots,n\}$	Energy consumed in SR segments

Chapter 3. Establishing the need for rigour

Growth in network energy consumption is widely cited in research that has the objective of controlling this consumption or improving the efficiency of consumption of energy to operate the network. Growth in the metro-core segment is predicted to grow at the fastest rate out of all segments [46]. To manage this growth, reliable reporting about actual and predicted consumption is fundamental.

Estimates of energy consumption in large telecommunications networks are available [14], [40], [47], [105]. The availability of several sources should serve to improve identification of the profile of energy consumption by segment of the network. While estimates will differ, it is at least expected that for large networks, the profiles would be comparable. It is not expected that major discrepancies arise when comparing the weight of any particular segment among the various sources.

However: one noteworthy doubt in this regard has in fact been raised. In [106], a contrast is made between the claim [14] that the network core will consume as much power (40%) as the access segment by 2017 and another claim [47] that the "metro/transport and core networks account only for 5 per-cent" in the same period (2015-2020). This doubt is resolved in this Chapter [13].

3.1 The initial impetus: standardize segment boundaries

The difficulty in comparing results of works that set out to assess energy consumption in the global network is well known [11], [40], [107]. A first solution is sought in two major steps. First, an organization that attempts to reconcile the boundaries of telecommunications networks (as defined in various works), is presented. Second the energy consumption projections (where common reference years may be found) are partitioned along these reconciled boundaries, and thus, the projections are compared. In so doing, highest common factors are identified, thereby establishing a base upon which dependent research may be grounded.

The effort presented in this chapter is described as an "initial" one, as it represents the first stage of an organic development of penetration into a dense and complex framework that is replete with technological and architectural variants. It is an attempt to organize that which is easily apparent in network deployments, using such terms as are readily available in first readings. This comparatively rudimentary modelling approach serves an ulterior need: to develop the powerful grasp afforded by acculturation. It opens new horizons into the framework, rendering visible that which was hidden, and comprehensible that which was opaque.

The rest of this chapter is organised as follows:

- Section 3.2 creates the case for harmonized reporting by cross-comparing members of a sample of the energy literature, in their use of segmentation and terminology;
- Section 3.3 suggests some premises in terminology for common elements within the Internet's architecture;

- Section 3.4 suggests a universal (albeit initial) method for segmentation of the Internet, and

Section 3.5 applies this method to compare the results presented in the chosen sample of the energy literature.

3.2 The case for harmonized reporting

This section shows the architecture of the broadband network found in a sample of three works from the energy literature. The three models are used here to make the case for harmonized reporting.

3.2.1 Ishii et al. [40]

Fig. 7(b) [40] shows the architecture underlying Ishii's work. This purports to be a representation of the structure of the broadband network that distributes the Internet in Japan. The segments are presented in bullet form for terseness.

- Access: This comprises the passive optical network (PON). It is rooted in a number of OLTs within the network operator's real estate and terminated within customers' real estate in an optical network unit (ONU).
- Aggregation: A ring of switches aggregates/distributes traffic within a zone of a metro area. A number of these rings cover the metro area.
- Metro-core: The edge router represents the IP routing function serving an administrative district of Japan known as a prefecture. The metro-core therefore comprises the switching boundary at which IP traffic is either switched to a different metro area within the prefecture or switched to the core network.

Core: This segment consists of the IP routers that comprise the distribution backbone of the Internet in Japan. Each core router may either switch traffic between edge routers that have a transport connection to it or between an edge router and another core router.

3.2.2 Bolla et al. [1]

Fig. 7 (d) [1] is described as a "typical access, metro and core device" network; the legend shows an access node, a transport node and a core node. This architecture is referred to in forecasts of energy consumption in Telecom Italia's broadband network [47].

- Access: This comprises a set of rings (blue), each of which is the logical topology of the interconnection between "access nodes". The access node is directly connected to customers. Customers' equipment is not shown in Fig. 7(d).
- Transport: A second set of rings (red) is shown. The caption to this figure [1] refers to "access and metro/core networks" and this same work refers to "transport network nodes".

Some equivalence can be deduced between the authors' intentions when referring to "metro" and "transport" segments.

- Core: This segment comprises the inter-metro backbone.

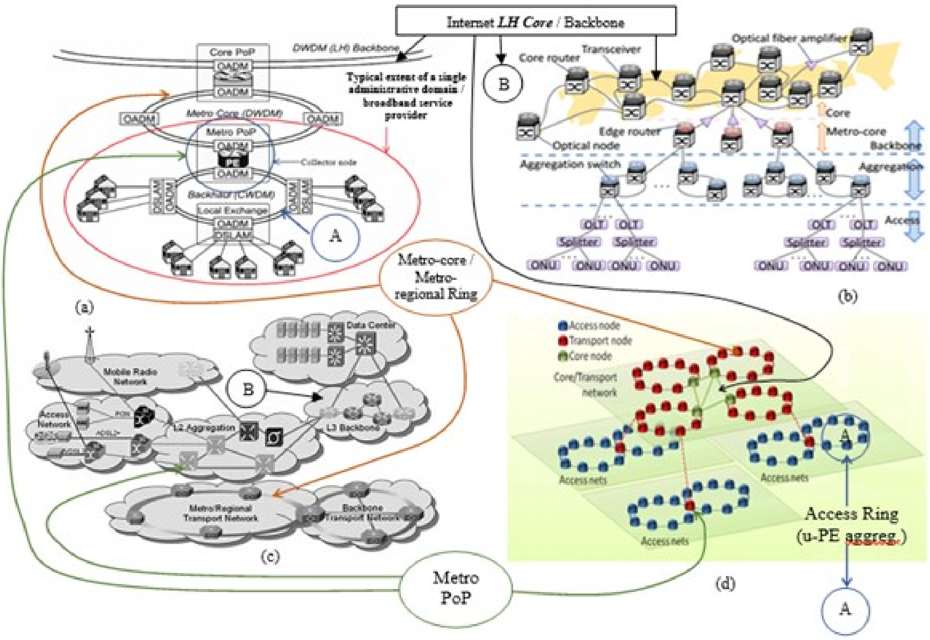


Fig. 7. **Reconciling sources describing the architecture of Internet network infrastructure (a) – (d) [1], [14], [40], [108].**

3.2.3 Lange et al. [14]

Fig. 7(c) [14] shows an "operator's broadband telecommunication network sections."

- Access Network: Several access technologies are included within this segment. Equipment in customers' premises is not included.
- L2 (layer 2) Aggregation: A tree-type, logical layer 2 network is described. The layer 1 hardware is referred to as a "metro/regional" transport network and described as an "underlying optical transport network (OTN)" ring.
- L3 Backbone: This is described as a partly-meshed backbone of IP/MPLS routers, overlying an OTN.

Even at this limited depth of investigation, the summary reveals several differences.

- Ishii includes customer equipment within the access section; Bolla and Lange do not.
- "Aggregation" is used by both Ishii and Lange but not by Bolla.
- Ishii divides the backbone into a metro-core and core. Bolla and Lange do not.

A harmonization of the boundaries is warranted to facilitate cross-comparison between reports. The harmonization must include a clear and sufficiently granular analysis of the segments of the network, to justify a fair analysis of these (and other) reports' conclusions.

3.3 Terminology

3.3.1 Metro-area architecture

IETF RFCs such as RFC 4364 [109] and RFC 4761 [110], as well as the Metro Ethernet Forum's (MEF's) architectural framework [111, Sec. 2], [111, App. II][6] are examples of standards that employ a consistent terminology to describe components of the metro-area network's infrastructure. The formal basis is established in RFC 4026 [20] (e.g., definition of Customer Edge device – CE, Provider Edge device – PE and provider routers that are not attached to CEs – "P routers"), which explicitly addresses the lexical difficulties that arose as provider-provisioned virtual private networks (PPVPNs) were investigated by several re-search groups. This terminology has been expanded by other RFCs, such as RFC 4761, which defines the user-facing Provider Edge device (u-PE). This scope of application of this terminology has expanded beyond the original scope of PPVPNs into the broader architecture of the Inter-net. Where possible, similar terminology will be applied here.

In [112], Herzog uses the term "transport" in a manner that requires the attention of a reader more accustomed to its use as the name of layer 4 of the OSI model. Indeed, use of "transport" in [112] (and by telco personnel in general), refers to the bit-pipe infrastructure: the transponders, muxponders, multiplexers, transceivers, cables, amplifiers, roadside cabinets, ducts, poles, real estate and other such elements that form the physical basis through which telecommunication is guided en route from one end to the other.

3.3.2 Providers

The term "provider" is now a hypernym for organizations characterised by diverse business models. Common labels include "telco", "carrier", "Communications Service Provider" (CSP), "Network Service Provider" (NSP), "ISP", "content provider", "telecom operator", "network operator", "access provider", "telecoms service provider", "public telecommunications network operator (PTNO)" and "telecom vendor".

Herzog [112] provides a good rationalization of the historical development of business models. The telco/carrier/telecom operator/network operator/access provider/telecoms service provider has (historically, at least) built and operated the network within the metro area and beyond it. Herzog reflects on the separation between networks and overlying services that is likely to characterize future business models. In [112], the term "telecom vendor" is used to describe the role that (a) provides connectivity at the physical layer in the form of shared (multi-tenant) infrastructure or dedicated infrastructure (e.g., dark fibre between endpoints), and (b) provides connectivity at the link layer in the form of virtual private networks (e.g., Metro Ethernet over WDM). It is the role that is commonly occupied by the incumbents, i.e., the organizations that have, traditionally: (a) dug trenches, laid ducts, erected poles, laid and strung cables and built offices to concentrate wiring and house switching nodes (b) designed link and node capacities to meet anticipated traffic and (c) operated, administered and managed the networks to ensure their stability. However, the term "telecom vendor" is not recognized widely enough to warrant its use as the name for the role. Indeed, none of the terms used in the first paragraph of this sub-section are unambiguous: they are only useful in the context of a broader text, speech or discussion. This problem precludes the instant recognition of the exact meaning of any of these terms.

The problem has been addressed in the Open Access Seamless Evolution (OASE) project [113], through stratification of the ingredients of a telecommunications service. The stratification is shown in Fig. 8 [113]. Henceforth, the term "CSP" will be used to refer the vertically integrated operator (the extent of the reach of the black parts), shown in cases (a) – (c). If necessary: if a subset of the role's functions is intended, then this will be specified. One dimension of the variety (shown in cases (a) – (c)) that inheres to this term (CSP) may be perceived through a description of the ownership of capital goods. Since such a description is necessarily heavily loaded with references to segments of the metro-area network, a limited elaboration is delegated to the sub-sub-section titled "Physical: real estate and topology", in the context of one of the segments.

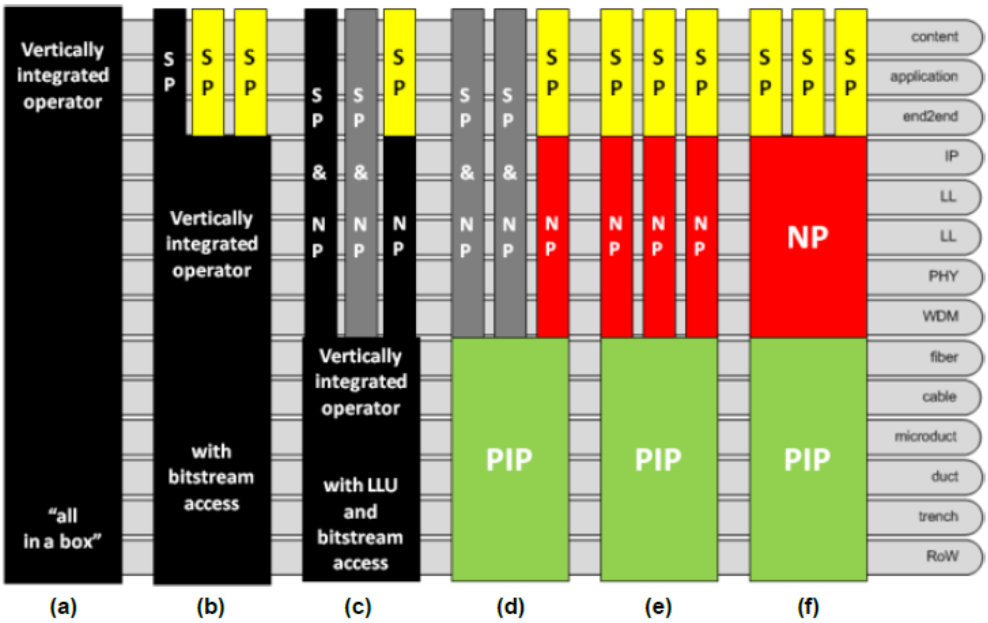


Fig. 8. Stratification of ingredients of a telecommunications service ("conceptual business models" [113])

The content provider corresponds to the yellow vertical bars in Fig. 8. The ISP is most likely similarly limited (i.e., yellow bars grouping), but it is possible that an ISP might operate the IP stratum too. Following Herzog's classification, the ISP and the content provider are either wholesale customers of the CSP or are part of the CS'' service set (this latter case is the vertically integrated operator one).

Therefore, the OASE stratification and Herzog's classification are in good agreement (although they use different terms).

The ISP and the content provider also consume datacentre infrastructure provided by "Internet Exchange Providers", who may be "carrier-neutral" or be part of a "carrier's" set of services. The term "carrier" is used here as this is the familiar one in IX parlance; however, a choice consistent with Herzog's classification would substitute "carrier" by "telecom vendor"; following the OASE, carrier would be substituted by vertically-integrated operator. This datacentre infrastructure is a point of convergence for interconnection (a) between peering CSPs (i.e., the interconnecting peers are all CSPs), (b) between peering ISPs (i.e., the interconnecting peers are all ISPs) and (c) between any combination of content providers (e.g., video library providers), CSPs and ISPs (the most liberal sense of peering). The term "Internet Exchange", commonly abbreviated as IX, is58odellint form of what used to be called the Network Access Point (NAP).

3.4 A first model: organization of Internet network infrastructure

3.4.1 The access segment

The access segment is the extent of the network that spans from the subscriber's premises to the provider's premises known variously as a Distribution Hub, Local Exchange (LE) – the latter may be referred to as a Central Office (CO), depending on the geographical region (e.g., CO is used in North America, while LE is used in the United Kingdom). Between the two end points, an important intermediate point in the architecture and distribution is the Remote Node (RN).

The active equipment that terminates this segment at the customer's end is commonly referred to as the Customer Premises Equipment (CPE). Within the terminology framework loosely identified in Section 3.3, the active equipment is referred to as the Customer Edge device (CE). At the Distribution Hub/LE (CO), a user-access convergence device (also known as: user-facing provider edge – u-PE) terminates the link. Examples of fixed-access (i.e., wireline) u-PEs include the DSLAM, the CMTS and the OLT). Wireless access convergence devices (such as the eNodeB and the gNodeB) add complexity, as they may link to a u-PE such an OLT, or they may function more like u-PEs in their own right. Moreover, of course, they are closer to subscribers and more numerous than distribution hubs, lEs or cOs.

The RN is located at kerbs and pavements, where it may be housed in a floor-mounted cabinet, in an enclosure on a pole or inside a manhole. It may serve as a demarcation point in the access segment; for e.g., in [70], the access segment is divided into a Secondary Access part and a Primary Access part, with the RN dividing the two parts.

1. Secondary Access network technologies include wireline PON, VDSL, DOCSIS and LTE. Secondary Access is commonly referred to as the last mile (or, conversely, as the

first mile, from the customer's perspective) and, as indicated, is demarcated at one end by subscribers' premises and at the other end by a roadside cabinet or pole-mounted enclosure. The components of Secondary Access may be found at various locations along the last mile, starting at the subscriber's end, proceeding through pathways towards roadside cabinets and roadside pole enclosures. The subscriber's end houses CPEs such as ONUs and CMs. The pathways include cabling ducts and pole-spans (overhead). The contents of the RN depend upon the mix-and-match of technologies that comprise the access segment. The roadside cabinet/enclosure may either host a u-PE, or it may host equipment that carries out the function of splitting the medium to serve a collection of cable runs to subscribers' premises. An example of the former would be a VDSL2 DSLAM; examples of the latter would be, respectively, (a) a GPON splitter, (b) a DOCSIS HFC optical node and (c) patch panels in ADSL/2 networks. Note that the latter grouping of functions is transparent even to OSI physical layer 1. The term Layer 0 may be used to refer to such functions. For example, splitting is commonly employed in GPON's ***optical distribution network***.

2. The Primary Access part spans from the cabinet/enclosure (wireline) or eNodeB/gNodeB site (wireless) to the access network operator's LE. The Primary Access part's technology stack may either be the same as the Secondary Access part's technology stack or it may be independent of it.
    a. Same technology stack: CMTSs and DSLAMs (ADSL) reach from the Distribution Hub/LE, all the way to the customer premises. In such cases, the RN would simply contain passive equipment or active equipment at layer 0
    b. Independent technology stack: VDSL2 DSLAMs may link to Metro Ethernet aggregation switches over LX/LH or ZX GE. Radio access network (RAN) eNodeBs may link (for backhaul purposes – see next paragraph) over PONs like ITU-T G.984 (GPON) or IEEE 802.3ah (GE-PON). The upper boundary of the access segment lies at the network-facing interface of the u-PE device.

A quick digression to establish the meaning of "backhaul" is warranted. This term is commonly used with reference to aggregation of individual subscribers' traffic on the access segment. The common interpretation of backhaul considers this aggregation to proceed as far as the boundary with the metro-core segment (see Section 3.4.3) of the network. This boundary is demarcated within the real estate housing the "Metro PoP".

The access network operator may sell services directly to subscribers (retail) or to service providers (wholesale) who do not have an access network in that geographical region, or sell both retail and wholesale. The type of service purchased by the subscriber is commonly referred to as "Broadband

Service”. Such a service may consist of a mix of best-effort service (e.g., residential Internet service) and service with service-level objectives (SLOs, e.g., VoIP and IPTV).

1) Deviations: demarcation of the access segment

One major disagreement in demarcation of the access segment in literature, regards the collector ring that physically interconnects the Distribution Hubs/LEs. In [108], this ring is considered as part of the backhaul network. There is no disagreement perceptible in this but this source proceeds to denote this ring as a metro access/backhaul, implying that the collector ring is part of the access segment. II-VI (an equipment manufacturer) is in tacit agreement: reference is made to a “metro-access ring” [114]. In [115], the same source (acting under the name of Finisar, which was acquired by II-VI) clarifies its understanding of the extent of the access network by graphically mapping it out in the context of a global network. It is a re-affirmation of a notion of the access segment as one that extends beyond the confines of the first major section of real estate, such as the LE or the Distribution Hub. Further affirmation of this understanding is found in [116] (**Error! Reference source not found.**). This view o f the network architecture is illustrated in **Error! Reference source not found.** [108], showing a hierarchy of rings, ending at the metro-core’s (see sub-section 3.4.3) boundary with the Internet’s core. Another deployment of a metro-area network is shown in **Error! Reference source not found.** [108].

2) Deviations: Demarcation of Primary and Secondary Access

Some access network distributions do not fit cleanly into the primary access – secondary access partition scheme. PONs that include large residential units (like apartment blocks) and large enterprises within their geographical reach may deploy splitters within the building complex or within a private services facility. In such a case, there is no intermediate demarcation between the Distribution Hub/LE and the customer’s premises.

Point-to-point (P2P) optical networks do not manifest a partitioning of the access network. Cables run from a local office, which may be no larger than a shed, directly towards customers’ premises. There is not even a user-aggregation device in such P2P access networks. These access networks may be complemented by PON deployments, to reflect product strategy.

The partitioning scheme is also complicated by next-generation optical networks that reduce or eliminate the need for distribution from the LE by exploiting long-reach optical technology to distribute directly from real estate that is higher in the distribution hierarchy. The position of the Optical Line Terminal changes in these next generation networks. Whereas current generation OLTs for GPON and GE-PON distribute primary feeder fibre cables from lEs and reach the kerb or the home, next-generation OLTs for WDM-PONs distribute fibre over longer distances, from. See Fig. 12 [108].

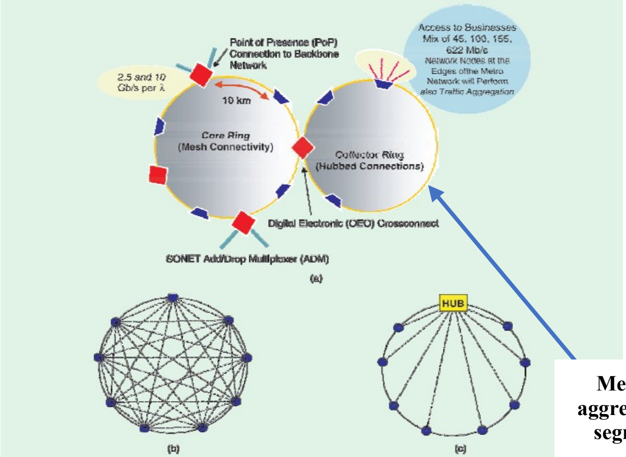


Fig. 9. The extended access segment: metro access ring is on the right [116].

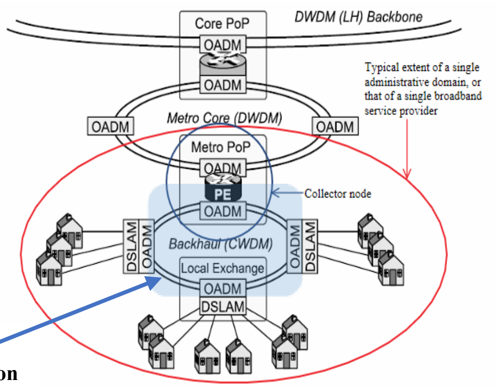


Fig. 10. Another view of a deployment of a metro-area network (from [108], adapted).

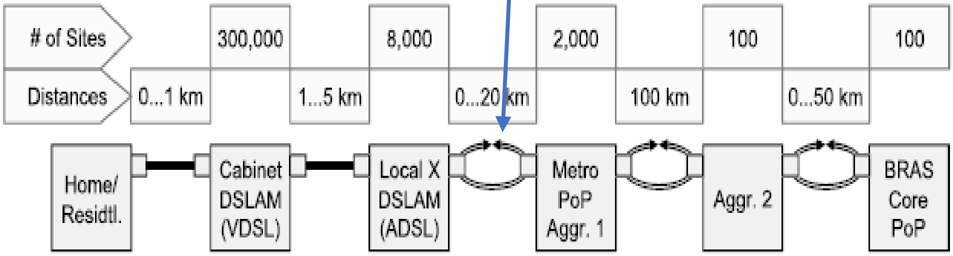


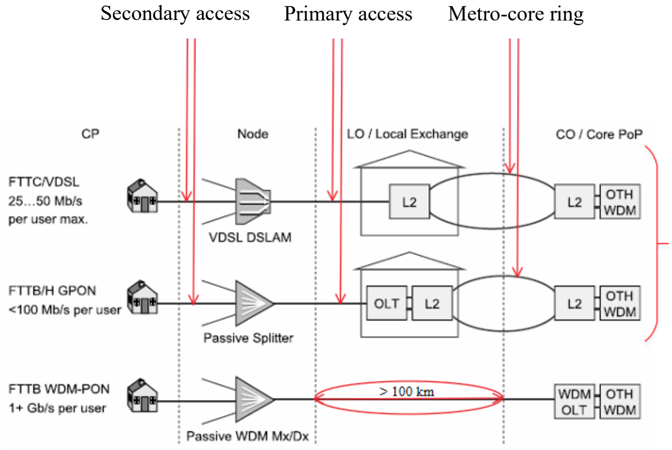
Fig. 11. The metro-area network, showing VDSL and ADSL in the “last mile”.

> **The structure shows various layers of ring in the distribution. The rings are typically optical networks that use either TDM (either current generation OTN (optical transport network), or legacy SONET/SDH) or WDM. Packet aggregation or TDM may be applied to improve WDM channel utilisation.** [108]

3.4.3 The metro-aggregation segment

1) The proposed boundaries

Proceeding upstream from the access segment, the metro area network commonly comprises a set of u-PEs (aggregation devices) and one or more Provider Edge aggregation switches. The u-PEs are housed in lEs that cover a CSP’s Service Delivery Area. The lEs’ traffic is back-hauled over a collector ring to the Metro PoP housing one or more aggregation switches (**Error! Reference source not found.**). The u-PEs include devices like ITU-T G.984.x / 1GE / 10GE OLTs, DOCSIS/EuroDOCSIS CMTS and DSLAMs. The aggregation switches include Metro Ethernet switches that aggregate traffic from several u-PE Layer 2 devices. These constitute the means of aggregation of the traffic of a number of access network divisions.



Simplification: metro PoP and metro access ring are omitted

Fig. 12. Division of the access segment changes with long reach passive optical networks [108]

The PE aggregation switches and the u-PEs minimally function as L2 devices but may also have limited L3 functionality [110]. The distribution of service edges outside the core and into the metro area of the Internet, in efforts to reduce energy consumption and improve QoS, creates new use cases for L3 connectivity between the CE devices and this segment. An illustration of the role of the aggregation switch is shown in Cisco marketing literature [117]. A comparison with **Error! Reference source not found.** shows good agreement between these two sources' segmentation of the metro area network, despite differences in terminology (arising out of the different perspectives from which these illustrations were drawn). The lower boundary of the metro-aggregation segment lies at the interface between the u-PEs and the aggregation switches.

The aggregation switches are themselves commonly interconnected in a ring topology (see Figs. 11,12) to two or more "Edge Routers" (PEs); the PEs are housed in Metro PoP real estate. The upper boundary of the metro-aggregation segment lies at the interface between the aggregation switch and the transport ring on which the PE router also has an interface.

The bases of the indicated choice of boundaries are two. Firstly, the partitioning is congruous with the intended applications of the technologies referred to. Secondly, a number of works have partitioned in a manner that bears a reasonable similarity to that described hitherto. Fig. 7 (introduced in Section 3.2) cross-references some of these works, using the segment labels that are proposed here. Fig. 7(b) [40] refers to an "aggregation" segment; this segment matches my use of "metro-aggregation" well. Fig. 7(c) [14] makes practically identical use of the term.

The illustrations included in Fig. 7 do manifest some deviations from the reference architecture which are being sketched in this chapter. For example, consider Fig. 7(d) [1]. There is no reference to an aggregation segment, yet inspection of the underlying work reveals that this is the collector ring gathering traffic from the u-PEs. This ring, therefore, is the ring of lEs.

The Metro PoP may also contain the boundary of a local broadband service provider's network. This would be the case where the broadband SP does not own the physical infrastructure downstream, but is only renting it (or part thereof) from an access provider. In this case, the Metro PoP may contain the broadband SP's PE routers used in the provision of Virtual Private Line Service (E-Line) and Virtual Private LAN Service (E-LAN).

2) An unfortunate choice of terminology: "aggregation"

The term "aggregation" has been used with reference to collection of traffic from subscribers by u-PEs, collection of traffic from u-PEs by PE aggregation switches and may be used to refer to collection of traffic from PE aggregation switches into another stage of link-layer aggregation switches (see "Aggr. 2" in **Error! Reference source not found.**). The term "backhaul" is also used to refer to t his act of collection of traffic from multiple L2 links onto fewer links having a higher bandwidth than those "lower" in the hierarchy. "Backhaul" is also interpreted diversely, with some definitions applying this as far back as the core of the network. See, for example, the note in [118, p. 4].

Some sources dispense entirely with references to the metro-aggregation segment (see Figs. 10 – 13). Another source [48] includes the segment in its description of the metro-area network, yet its boundaries lack crisp definition. A publication complementary to this source [105] manifests the same blur. Two distinct segments – "Access and Aggregation" and "Metro" – are presented. The term "Access and Aggregation" is itself problematic and no substantial justification is given for the choice of words. The description of what comprises the "Metro" segment compares well with the contents of the metro-aggregation segment, despite the lack of architectural detail. Indeed, [70] refers to a "Metro" segment and the description given also compares well with the metro-aggregation segment.

Summarizing: this segment has been identified by no less than the following names: "metro access" [108], [114]; "backhaul" [108], or part thereof (as indicated by [118]); part of "access" [1], [115]; "collector" [116], [119, p. 153] and "metro collector" [119, p. 170]; "metro" [105], and last (but not least) "metro-aggregation" [40], [48]. Metro-aggregation is comparatively unambiguous by virtue of its breadth of use. Further on in this chapter, a description of its contents is presented, in a manner that facilitates classification of technologies and minor architectural variations.

3.4.4 The metro-core segment

1) Physical: real estate and topology

The metro-core segment connects a number of Metro PoPs and one or more Core PoPs per metro area. Physical topology of interconnection is commonly a ring [108], [116], [119, p. 145,152], [120], [121, p. 157][8] [11] [21] [22, p.145, p.152] [24, p.157]. For example: a DWDM (dense wavelength division multiplexing) ring, installed in 1+1 redundancy (traffic + identical copy of traffic both flowing concurrently) for protection, may link the Metro PoPs to the Core PoP(s)[108]. An illustration of such a topology is shown in **Error! Reference source not found.**, where a metro core D

WDM 1+1 ring is shown in the context of a metro-area deployment. The metro-core is also referred to as the metro-regional segment.

The brief elaboration on CSP diversity, referred to earlier, can now follow. Ownership (by the CSP) of the capital goods comprising the segment varies across a range of consolidation.

- At one end, all such goods might be owned by a single CSP. The CSP would own the metro-core ring transport layer (OSI Layer 1) hardware as well as the premises hosting the Metro PoPs and the Core PoP. Such is the case of Telecom Italia's (a good example of an incumbent) metro-area network in several Italian cities, interconnecting DSLAMs at the access end to the national backbone at the long-haul core end [108, p. 360]. This end of the range corresponds to the cases of the vertically integrated telecommunications provider. Such an operator would occupy Open Access network scenarios shown in Fig. 8(a) – (c) [113].
- At the other end of the range, ownership is highly fragmented. One operator would own the transport ring hardware. The Metro PoPs might be located in carrier-neutral exchanges / colocation centres, where network providers connect to their wholesale clients like ISPs. The Core PoP may be hosted in a carrier-neutral data centre serving as an Internet Exchange for the ISPs in the metro area [122]. Open Access design drives ownership distribution towards the interoperability symbolized in the case shown in Fig. 8(f) [113]. Case (f) represents this end.

The metro-core segment is the network that links the carriers that cover the same metro area. It is also the network that interfaces to both the metro-aggregation segment and the core segment. The metro-aggregation segment appears as several "metro-edge" (another term!) rings that are interconnected with the backbone network in Fig. 13 [121, p. 158].

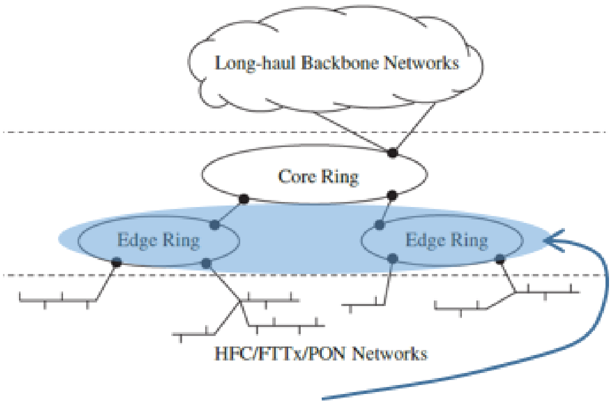


Fig. 13. Multiple rings in the metro-aggregation segment connect to the metro-core ring [121, p. 158]

Smaller metro areas served by very few CSPs may not have a metro core ring and a Core PoP at all. For example, for the sake of Internet traffic exchange, these CSPs might peer directly. The physical location may perhaps consist of real estate adjunct to one of the peers' hosting locations. Such an arrangement establishes peering connectivity without granting reciprocal access to premises hosting closely guarded infrastructure.

2) Logical: Traffic flow

This segment accumulates traffic from the provider(s)'s points of presence within the metro area (Metro PoPs); conversely, it distributes traffic to these Metro PoPs. Traffic flows vertically between any Metro PoP and any Core PoP. Fig. 14 [19] illustrates the (logical) relationship be-tween individual ISPs' Metro PoPs and the Core PoP. Traffic also flows between Metro PoPs (not shown).

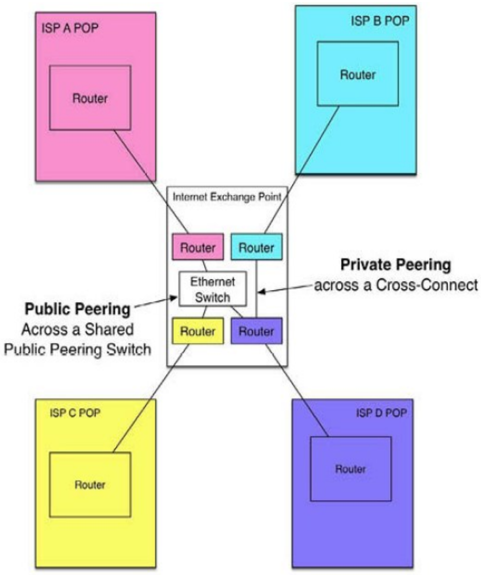


Fig. 14. The metro-core segment comprises a number of Metro PoPs that are logically interconnected at an Internet Exchange that also serves as a Core PoP [19]

The flow is characterised as meshed [116], [119, p. 148] [22, p.148]; see **Error! Reference source not found.**, the part labelled (b), on the bottom left. The meshing is accomplished through the use of optical add-drop multiplexers (OADMs) at each node of the ring. By passing through an OADM, a lightpath of a given wavelength renders the node transparent and forms a logical connection between the node of insertion (add) and the node of removal (drop). The upper boundary of the metro-core segment lies inside the Core PoP, at the transport interface(s) between the P-routers and the metro-core fibre. This explicitly excludes Core PoP P-routers from the metro-core segment and establishes their transport interface to the metro-core fibre as the boundary between the metro-core and the long-haul (LH) core.

3) Functionality

The traditional functionality of this segment has been twofold. One function is that of extending the geographical reach of the network to cover longer distances than those possible with the technologies used in the access and metro-aggregation segments. The segment bridges the access and aggregation segments to the long-haul backbone network [119, p. 152]. The logic of the division of reach includes the important purpose of reduction of network node complexity. Nodes on shorter links have fewer functional requirements and are less costly to deploy and operate. The second function is that of IP routing. The purpose here is not to delve into the relationship between IP as a client of an underlying transport layer such as DWDM with OADMs at the nodes. The purpose is to identify this layer as that in which IP routing between intra-metro endpoints takes place.

The term metro-core is changing under the pressures of traffic growth [46], and this leads to difficulty in reconciling some works with others. As presented thus far, the metro-core segment may be viewed as a segment that aggregates/distributes traffic between the long-haul backbone segment and the metro-aggregation segment (inter-metro), as well as routing traffic within the metro. This is not universally true. Some metro networks have limited or no intra-metro switching capability [123]. As late as 2009, direct reference to the routing function is omitted from the metro network in a well-cited work [14]. In that work, the segment's function seems to be included within the "L3 backbone" segment as there is no reference to the L3 function outside that. In an Alcatel-Lucent TechZine article [124], arguments are made in favour of "introducing a metro core into the metro aggregation network". The benefits identified may be summarized as follows: reduction of the length of the path between source and destination. Since traffic flows now increasingly have a source and destination within the metro area, then a routing core capable of switching all such traffic should be part of the metro area network. Fig. 15 [124] shows the stage of insertion of the routing. Fig. 16 [46] shows the location within the broader context of the metro area network. This graphic amply demonstrates the difference between the view that delegates the routing function to the long-haul backbone [14], [46], [124] and the view that includes it within the metro-core [116], [119]. In the former, the metro-core does not exist as a separate segment; in the latter, it is a segment that affords meshed logical connectivity albeit over a physical ring topology. The view in which the metro-core does not exist as a separate segment but rather is integrated within the backbone will be referred to as the first view. The second view, conversely, is that which considers the metro-core as a segment that supports richly distributed (meshed) connectivity between its nodes.

The metro-core router in Fig. 16 [46] seems to have a strikingly similar role to that of the group of routers shown in Fig. 14 [19] inside the Internet Exchange. The resemblance is not coincidental; their roles are indeed similar. The difference lies in the consolidation implicit in the ownership of the architecture. Both the sources (quoted earlier, i.e., [14], [46]) that seem to ignore the existence of a separate metro-core segment relate to vertically integrated operators, whereas Fig. 14 is clearly exhibiting higher degrees of openness according to the Open Access Network set of scenarios. In the circumstance of the vertically-integrated operator, both these sources ([14], [46]) classify the metro-core router as part of the metro area infrastructure, but from the perspective of the ISPs in the multi-player ecosystem shown in Fig. 14, the metro presence ends at the Metro PoP. This rationale is

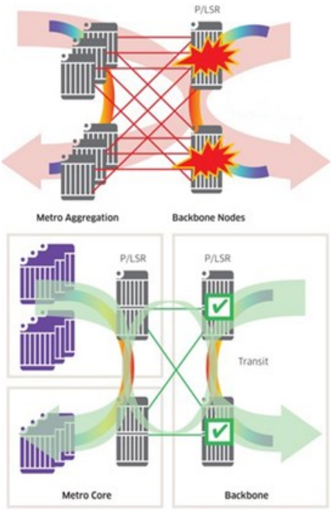


Fig. 15. **The metro-core segment introduces routing functionality between the metro-aggregation and Internet backbone [124].**

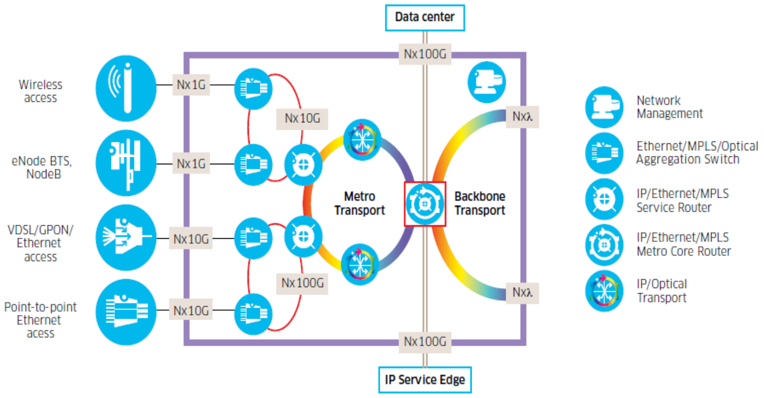


Fig. 16. **Enhanced routing functionality inserted here to prevent traffic from unnecessarily transiting to the backbone [46].**

confirmed [46] by consideration of the use of a particular integrated-services model of metro-core router in an Internet Exchange application as beyond the scope of a metro-core deployment.

The switching of intra-metro endpoints' traffic away from the long-haul core segment may be thought of as a functional description of the metro-core segment. It is achieved through the insertion of routing hardware between the metro-aggregation and long-haul core segments. Note that the functional description of the metro-core segment shows that notwithstanding the absence of a Core PoP, smaller metro areas can still benefit by establishing the functionality of this segment.

3.4.5 Result of the initial impetus

This section (section 3.4) has been written to bring the system boundaries into sharper relief, as they are essential to a good understanding of trends in energy consumption associated with the transmission, transport, switching and routing of traffic. The section concludes here, with a digest in the form of a graphic (Fig. 17) that attempts to facilitate understanding of this section. The illustration in Fig. 17 is an essential construct in this initial impetus towards standardization of segment boundaries. Through the use of representative technologies and topologies, it shows:

1. the principal segments
   a. access,
   b. metro-aggregation and
   c. metro-core and
2. the system boundaries.

Henceforth, an assessment of the strengths and weaknesses of these initial results will be used as the basis for development of a proposal for a draft of a standardized representation of metro-area networks.

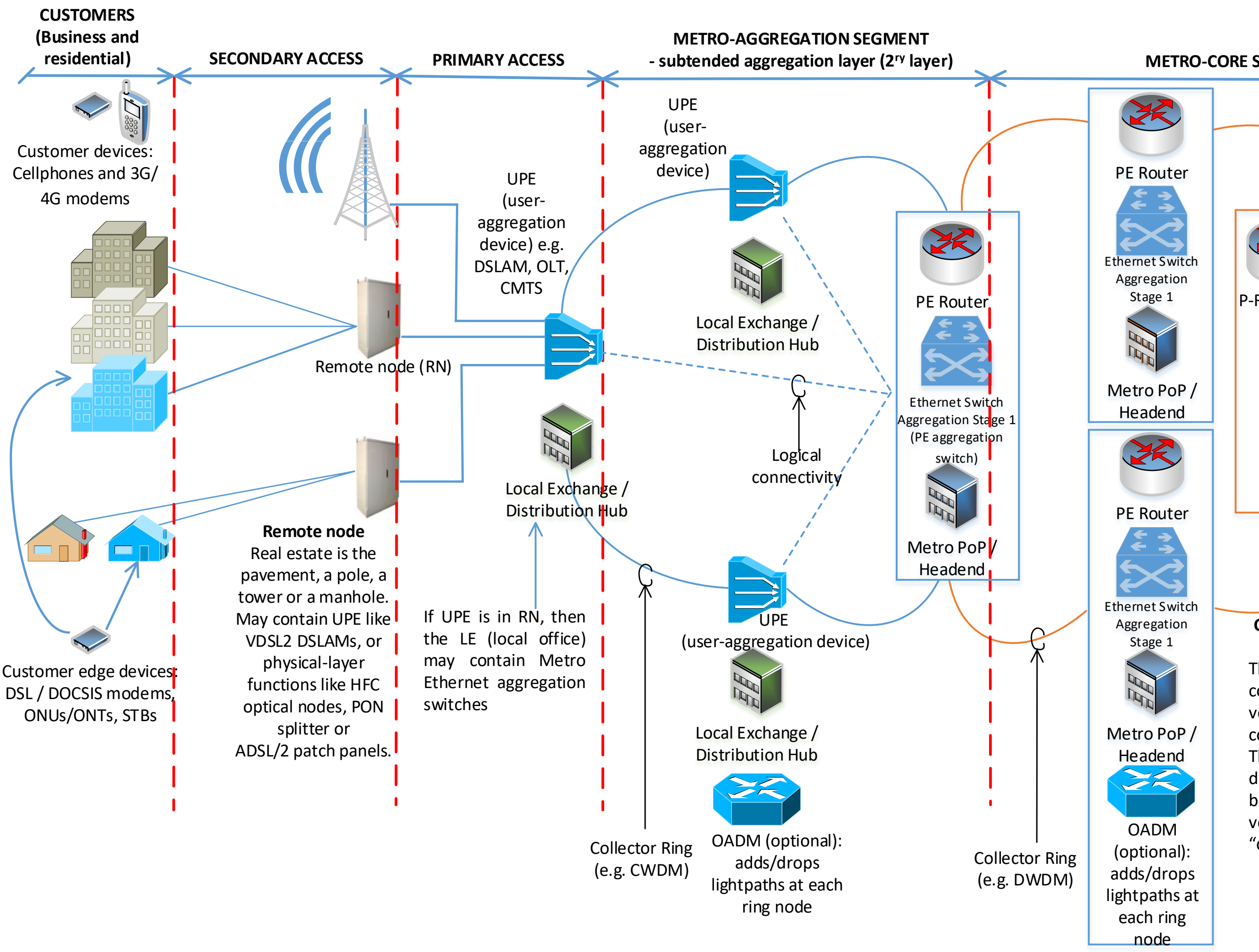


Fig. 17. Illustration complementing the recommended model of segmentation of the metro-area network for analysis and reporting of energy consumption

3.5 Comparison of four reports of Internet energy consumption

This section first analyses the segmentation proposed by a sample of publications in the energy literature, then cross-compares their numerical results to the extent permitted by the identifiable alignment of the segments. A summary of the segmentation analysis is shown in Table III . Terms within braces are those used by the authors. A summary of the comparison of their numerical results (where relevant, as not all works estimate energy consumption totals) is shown in Table III and Table IV .

Table III COMPARISON OF SYSTEM BOUNDARIES WITH THE ADOPTED TERMINOLOGY

Ref. #	Access	Metro-Agg.	Metro-core	Long-haul Core
[40]	{Access}a	{Aggregation}	{Aggregation}	{Backbone}
[14]	{Access}	{Aggregation}	{Aggregation}	{Backbone}
[47]	{Access}	{Access}	{Metro/Transport}	{Core}
[69]	{Access}	{Metro}	{Metro}	{Core}

a. {Access} = ***access*** + ***CE***

3.5.1 Analysis of segmentation proposed in the energy literature

1) Internet in Japan [40]

The authors propose three primary segments: access, aggregation and backbone, the latter divided into metro-core and core.

a) Access = access + CE devices

The boundary is shown as the network-facing interface of the OLT. This boundary matches the definition proposed in II-B. However, the authors' use of "access" segment includes ONUs; therefore, the comparison identifies this mismatch.

b) Aggregation = metro-aggregation + metro-core

A single ring of aggregation switches is shown. This architecture matches that shown in Fig. 12. This figure shows that a simplification has been made: the Metro PoP and Metro Access rings (metro aggregation) have been omitted. Therefore, it is immediately visible that some compromise must be made to match this segment with one or other of metro-aggregation and metro-core. There is no clean fit. The logical position of the "edge router" indicated in Fig. 7(b) [40] does not fit that of the PE-router. The "edge router" links to other core routers using an erbium-doped fiber amplifier (EDFA)); a PE-router connects to a LH-core router intra-Core-PoP, without such amplification. The most reasonable match would be to place this "aggregation" segment in the same class as the metro-aggregation and metro-core segments proposed here and locate the "metro-core" inside the LH core. Since the metro-core has no layer 3 functionality in this case, it corresponds to the first view expressed in sub-section 3.4.4. Nippon Telephone and Telegraph (NTT) is the incumbent in Japan. Its ownership of capital goods compares well with that of a vertically-integrated operator; this strengthens the correlation between the first view and the vertically-integrated operator.

c) Backbone (core + metro-core) = Long-haul Core

This segment visibly corresponds to the long-haul core.

2) Internet in Germany: Deutsche Telekom [14]

a) Access segment = access

Various access networks are illustrated, e.g., VDSL2, ADSL2+ and PON. In each case, the network-facing side of the u-PE is the upper boundary of the segment. Therefore, this matches the definition proposed here (see sub-section 3.4.1) of the access segment.

b) L2 Aggregation = metro-aggregation + metro-core

A distinction is made between client layer and server layer: the optical transport network is depicted as the server for the aggregation technology chosen. Only one ring is shown but the label attached to it ("metro/regional"), as well as the evidently summative intention of the authors in illustrating this segment (reproduced in Fig. 7(c)) as the intermediary between the u-PEs and the "L3 backbone", leave little room for doubt that this segment best matches the joint metro-aggregation and metro-core segments.

As with NTT'S case, Deutsche Telekom is a vertically-integrated operator. This further strengthens the correlation between the first view and the vertically-integrated operator.

c) L3 Backbone = Long-haul Core

This segment corresponds to the long-haul core.

3) Internet in Italy: Telecom Italia [1]

The boundaries used in [47] are not immediately identifiable, as there is no explicit reference within the document to an architecture. Since it is implied that "traffic load values" used in this work are the same as those used in [1], this latter work was examined to extract an interpretation. The architecture is shown in Fig. 7(d) [1, Fig. 8]. Interpretation is not straightforward, as the iconography is basic.

a) Access nets = access + metro-aggregation

Comparison with **Error! Reference source not found.** assists in the identification of the access n ets as the rings that backhaul traffic from the u-PEs. Since the access segment is evidently essential in the metro area network, it is taken to be implicit in "access nets".

b) Transport network = metro-core

The hierarchical position of the nodes of the transport network, as well as their site at the intersection between two segments, identifies the transport network as the metro-core.

c) Core network = Long-haul core

This segment visibly corresponds to the long-haul core.

4) From the perspective of an early study on energy efficiency of video on demand services

Jayasundara et al. [69] investigated improvements in energy efficiency attainable by moving video caches closer to the point of consumption. While the model [69, Fig. 1] does not identify detail about physical topology, it includes sufficient information to justify a comparison with the segments presented here.

"Access" and "core" are readily identifiable with the access and long-haul core segments respectively. Inspection of the model's "metro" segment shows that despite the lack of detail about physical topology, there are two distinct parts to this segment. One part comprises a network between PE routers; the other part comprises an aggregation network that backhauls traffic from the "access" part of the model. Therefore, this model distinguishes between a metro-core and a metro-aggregation component but lumps them under the "metro" designation. Despite the superficial similarity with Ishii's edge router [40], Jayasundara's model distinguishes itself because it separates the P-router's function from the PE-router's function. Ishii et al. do not evidently distinguish between the long-haul core router, which interfaces to the LH backbone, and the PE router. The two seem to be lumped. This example illustrates the importance of distinguishing between logical and physical topologies in modelling. Cross-comparison of numerical results: 2017

The three studies [14], [40], [47] are compared for the year 2017, which is part of all three studies' estimates. Of the three, [47] estimates the energy consumption for a five-year period (2015-2020). A summary of the comparison is shown in Tables III and IV.

Table IV PERCENTAGE OF ENERGY CONSUMPTION FOUND IN THREE DIFFERENT STUDIES FOR YEAR 2017

<table>
<tr><th>Ref. #</th><th>CE</th><th>Access</th><th>Metro-Aggregation</th><th>Metro Core</th><th>LH Core</th></tr>
<tr><td>[40]</td><td colspan="2">72.8</td><td colspan="2">6.5</td><td>20.7</td></tr>
<tr><td>[14]</td><td colspan="2">86.3</td><td colspan="2">6.4</td><td>7.3</td></tr>
</table>

Table V PERCENTAGE OF ENERGY CONSUMPTION FOUND IN THREE DIFFERENT STUDIES FOR YEAR 2017

<table>
<tr><th>Ref. #</th><th>Access</th><th>Metro-Aggregation</th><th>Metro Core</th><th>LH Core</th></tr>
<tr><td>[14]</td><td>39.3</td><td colspan="2">21.4</td><td>39.3</td></tr>
<tr><td>[47]</td><td colspan="2">74.2</td><td>22.2</td><td>3.6</td></tr>
</table>

Each study has some differentiators that complicate direct comparison. For example, Japan's operators are planning to shut down use of DSL as fibre-to-the-home's (FTTH's) market share increasingly justifies it [125]. In 2013, Japan's fixed broad-band penetration rate (73%) into households was substantially higher than Germany's (64%) or Italy's (49%). Furthermore, in the study of Japan's Internet [40], ONUs are included in the access segment calculations; this aggregates the CE devices consumption inextricably into the access segment's estimate and precludes some comparison (e.g., with [47]).

Notwithstanding such difficulties, the noteworthy doubt [106] identified in the introduction to this Chapter, can be resolved, as follows:

- the figure of ***5% consumption*** [47] by the network emerges when this is taken relative to the total that includes CE devices, whereas
- the figure of 40% [14] excludes it. Indeed, if the CE devices are taken into account, the energy consumption of the LH core and metro-core in [14] is estimated to be ***between 7.3% and 13.7%*** in 2017.
- Conversely, if the energy consumption in the cEs is excluded (as in [1]), then the percentage of the energy consumption in the metro-core and core (according to the boundary estimations shown in Table IV) is $\frac{92+15}{92+15+307} \times 100\% ,= 25.8\%$.

Moreover, as indicated in [40], the impact of improvements in energy efficiency was not taken into account in [14].

3.6 A first recapitulation

Before delving into observations on strengths and weaknesses (Chapter 4), a first recapitulation is in order. This chapter carries recommendations in favour of the rationalization of reporting structure and terminology, to facilitate cross-comparison between future efforts at measuring and estimating the growth of energy consumption in the Internet. To this end, a model describing segments and system boundaries of the metro-area network, has been suggested (see sub-section 3.4.5 and Fig. 17). This model is part of the foundation upon which a more detailed reporting framework may be built and a standard developed. A case has been made for the value of these efforts, by drawing attention to the difficulty of cross-comparison where the segments of the architecture either do not include the same set of components, or the presence of specific components is ignored.

Such a rationalization may also be applied to other application domains. For example, there are various sources of traffic estimation, such as Cisco's Visual Networking Index, Sandvine's annual reports and Bell Labs' (now Nokia) publications (particularly in so far as concerns their involvement in GreenTouch). The perspectives of the reports vary. A rationalization of the various sources may be based upon the same work as that carried out to produce a standard for Internet reporting frameworks.

Chapter 4. Investigating the Implementational Model

Perhaps a simpler, and pithier recapitulation (at least of the maladies) than that given in the conclusion to the previous chapter, is the following. Anders S.G. Andrae unequivocally warns energy analysts of the peril of double counting of energy use due to " 'wrong' slicing " [126] (segmentation problem). William B. Norton, with similar bluntness, stresses that "The Lexicon Is Important" (terminology problem). Both these problems were stressed upon in Chapter 3.

- The inconsistent use of terms was commented upon and "aggregation" and "backhaul" were identified as two examples;
- it was proven that, without a reference model against which to compare energy use estimates, these estimates have the potential to mislead readers.

A justifiable observation on the work in Chapter 3 is that it may be insufficiently abstract to be directly usable as a model. This argument would fairly claim that to qualify as a model, a representation must satisfy the criterion of generality. Since the representation shown in that work included references to technologies and topological structures, it may be insufficiently general to qualify as a model. It follows that the effort to build upon that work must investigate public telecommunications networks without committing to excessive detail about implementation.

Therefore, a balance must be found between abstraction and implementation. At the sweet spot, the model will meet the need for generality while emphasizing the physical organization of the telecommunications network. ITU-T Y.2011 aptly describes this as a concern with the way in which "functions are distributed and implemented in equipment" [127, p. 8]. Within standardization documents regarding the ITU-T's Global Information Infrastructure (GII) (e.g., [100], [127]), this kind of modelling is described as "implementation model"-ing or "implementational model"-ing. This type of model avoids the technological and topological trappings that limit the use of the result shown in Chapter 3. This is a weakness of that model.

Efforts must also be guided by the needs of the intended user of this work: the use-phase energy analyst. This analyst undertakes the design and execution of a study that carries out some or all of the activities of measurement, reporting and analysis of the use of energy during the use phase of the life cycle of a telecommunications network. Ultimately, analysis of energy use cannot avoid resolution of the general form into some implementation. This, therefore, presents the strength of the previous model: it represents an instantiation, encompassing a variety of technologies, of a metro-area network. The requirement for instantiation characterizes the area of research, since energy use must be bound to physical entities. An attempt to bind energy use to logical entities can proceed with analysis only until it reaches the point where it has determined what the functional unit [91] consists of. After this, the energy used to produce the functional unit requires concretization into specific, physically-recognizable devices.

4.1 Objective

From this background, the profile of development of the implementational model emerges:

- domain ontologies are mined (in particular that of the ITU-T's GII [100]), to establish the foundations of the method (notably: terms and concepts), and therefrom,
- a reference model is suggested, suited to use as a template against which data collection and report analysis can be carried out.

The work in this chapter is thus framed by the ITU-T's principles and framework for the GII (recommendation Y.110, [100]). Recommendation Y.110 is intended to provide "input for … development of detailed standards for critical functionalities required to support the enhancement of the GII" [100, p. 1]. In pursuit of this goal, Y.110 describes the GII implementational model. An implementational model interests the energy analyst in so far as:

- it describes, at a high level, how functions of the telecommunications network take physical form;
- it groups functions into segments
    - to identify and emphasize such functions as are implemented as a group and
    - to identify and emphasize segment interfaces as good candidates for standardization.

The objective is, therefore, that of extending the energy literature through the provision of a reference model for the classification of energy use by segment of implementation.

The rest of this chapter is organized as follows.

- In Section 4.2, the method used is described in detail. The description will show how the method emerges from the general methodology prescribed in Y.110. Note the relationship between methodology and method implied here: Y.110 provides the means to develop detailed standards. It provides the methodology in the classical interpretation of the term, i.e., as a matrix out of which methods are developed.
- In Section 4.3, the method is applied to organize existing models; thereafter, the emergent organization is used to derive the implementational model. Analysis proceeds in ascending order of complexity of (these existing) models found in literature. Ordering uses stage numbers, with Stage 1 being the simplest and Stage n+1 being more complex than Stage n.
- In Section 4.4, guidance is suggested, intended to draw the energy analyst's attention to good practice, common pitfalls and fallacies. This guidance is a by-product of processing, obtained through the perspective afforded by the method as it proceeds through the collection, organization and analysis of existing models.

4.2 Method

4.2.1 Overview

The objective may be summarized as that of designing a model that abstracts the physical technologies but uses ***reference points*** (RPs) [128]; in particular: ***reference points for interconnection*** [101]. Reference points are used to separate and draw attention to (a) functional groups and (b) segments within the metro area. Here, segments are used as defined in Y.110 [100], but use of segments is made in a manner that is not limited to the implementational model suggested there [100, p. 32]. Analysis (see Section 4.3) shows that Y.110's segmentation is insufficient to fully characterize energy use.

For data, observed models are collected and reconciled to the extent permissible from the information which each model contains. The difference in detail between the models requires an organization of the models into levels of complexity. As soon as it becomes possible to resolve sufficient detail in the models to recognize reference points, the reference points are applied at the appropriate interconnection.

4.2.2 Which networks are within scope of modelling?

This sub-section employs terms as used (in some cases, even defined, albeit in a "semi-formal" manner) in Y.110 [100]. Scope is limited to that of telecommunications networks which are deployed to fulfil the "infrastructural role" of the "network operator". This latter term lies within the scope of the term CSP, but recommendation Y.110 is noteworthy in that it moves away from the vertically-integrated instantiation of the CSP and towards the concept of "value chain", where several roles complement one another in the delivery of services to end users. The term "infrastructural role" is also defined in Y.110 [100]: this role acts as a "supplier" for all value chains. The vision of disaggregated provision of service is explicit throughout Y.110.

On the other hand, this chapter's scope excludes networks owned by "structural role" players [100], such as content providers and application service providers. It also excludes infrastructural role players like Content Delivery Network (CDN) providers and Internet Service Providers. This is necessary in order to limit the size of the problem posed in the first thrust towards an implementational model.

Within this scope, the usefulness of this modelling work is further limited by the breadth of implementation of the telecommunications networks which it represents. For example, public switched telephone networks (PSTNs) commonly employ time-division multiplexed trunks between local exchanges. Similarly, cellular base stations may employ microwave links as trunks. Both these cases exemplify network separation, and stand in contrast with the packet trunks that are a foundational underpinning of converged networking. Note that the discrimination between converged networking and network separation does not inherently discriminate between best-effort and service guarantees. Converged networks have long been capable of traffic policing and traffic shaping. The distinction made here, rather, is between services that have more or less of the network in common with one another. The fewer the network commonalities between services, then the more abstract the model. However, the flip side poses a danger to the modeller: the more the commonalities, the easier it is to get ensnared in technologies and to tend towards concrete deployments in modelling.

One useful approach to determining the appropriate level of abstraction is to seek a high-level, comprehensive view, enriched by ***a sense of the momentum of developments of functional architectures***. This approach may be resolved into two investigations that seek to identify the functional architecture.

- Firstly, there is the direction given by Standards Developing Organizations (SDOs).
- Secondly, there is empirical evidence from the field.

Both investigations have been undertaken, and findings are described next.

1) SDOs' Direction

a) ITU-T

The ITU-T sees that the target of the next-generation network (NGN) is to ensure global interoperability and tasks itself with the development of "Recommendations, Standards and … implementation guidelines for the realization of [] Next Generation Networks." (Y.2001 [129]). The networks within scope are (a) packet-based, (b) capable of telecommunications at standard levels of QoS and (c) capable of layered services. Importantly, a fundamental characteristic is that of decoupling service (control and provision) from transport [129]. Y.2011 [127] (which is a standard complementary to Y.2001) accentuates even further, referring to this separation as a "key cornerstone" of NGN characteristics.

b) ETSI (European Telecommunications Standards Institute)

ETSI's standardization efforts in the space of functional architectures are compliant with those of the ITU-T, i.e., ETSI's standards and technical reports are developments founded on the ITU-T's work. For example, ETSI's NGN functional architecture standard, ES 282 001 [130], is compliant with the ITU-T's general reference model [127]. ETSI also seems to have a collaborative approach to its mission. For example, work on IP interconnection of voice over IP (VoIP) [131] seeks to fill a gap left by other organizations [131, p. 5]. Furthermore, the IP Multimedia Subsystem (IMS) is central to all services which ETSI standardizes. The 3GPP's work was adopted and developed. Work on the IP Multimedia Subsystem (IMS) was transferred back to the conceiving SDO (3GPP) with the declared objective of assuring a single source of standardization of this space [132].

c) Broadband Forum

The Broadband Forum's technical report on architecture and framework for broadband multi-service, TR-144 [133], includes a list of legacy applications that must be supported by the broadband multi-service network. These include the public switched telephone network (PSTN) and TDM circuits for PSTN, business services and mobile telephony backhaul. These are described in TR-145 [134] as "legacy access technologies". These two documents stress convergence onto a common (broadband) network to distinguish legacy services from NGN services.

d) OECD

While not strictly a standardization effort, OECD Digital Economy Paper #207 [135] fits better here than in the group showing empirical evidence. It provides further direction. The authors refer to "experience" that indicates that the packet-centric model will continue to displace the circuit-centric TDM model where it still has some limited hold. It is important to point out that the intention here is not to displace TDM's localized use. It is still useful for sub-wavelength, intra-segment networking. It is the broader use over concatenated segments that has been and continues to be displaced.

e) SDO recap

The concept of the NGN is elaborated by all the major SDOs reviewed. Core themes emerge.

(a) Network convergence is one. Broadband access networks are intended to replace service-specific networks (e.g., PSTN and SDH/SONET (Synchronous Optical Network/Synchronous Digital Hierarchy) networks). IP supports this convergence through the abstraction of the transport; through this same means, it also enables "nomadism"/ "generalized mobility" (another theme).

(b) Support of ("interworking") legacy services is another.

(c) Yet another is the complementary pair of open interfaces and open access, through which the network provider collaborates with business partners or is compelled to do so by regulation.

(d) Perhaps the theme which has greatest implications for the energy analyst's work is independence of service (provision and control) from transport. This theme is intertwined with that of open access and open interfaces but its implications are so variegated that it merits a separate reference.

2) Empirical evidence: the evolution of Video Distribution Networks

The current and projected dominance of video in IP traffic has been reported by Sandvine (Global Internet Phenomena report) and Cisco (Visual Networking Index, e.g., [44]). Cisco further distinguishes between managed IP video (delivered by a service provider over a single public telecommunications network (PTN), without crossing PTN boundaries) and Internet video (OTT). For both delivery models, the VNI reports current and projected dominance of video among IP traffic. Cisco's reference to managed IP video growth is just part of a narrative that describes the transition of video from an isolated suite of technologies to integration with IP, both managed and Internet-delivered.

A clear illustration of this statement is obtained through examination of video production in studios and headends.

Video production has evolved in a technological environment designed specifically for this application domain's requirement. A particular and pressing concern in video production has been sensitivity to frame-accurate timing to deliver tear-free video. This is especially important while switching content sources (for adverts and other brief inserts). It has therefore been necessary to mix-and-match video signals from different sources, live as well as recorded. The solutions that addressed these challenges were built upon the baseband transport of digital video signals (over serial digital interface (SDI)) and domain-specific digital video routers. The major downside is the architecturally-ingrained rigidity: each SDI cable carries a single video signal and the digital video router is a crosspoint-matrix that connects a single input – and therefore a single signal – to a single output.

Technical challenges to supporting the domain's requirement in an IP-based environment have been overcome, with one of the latest significant advances being the release in late 2017 of the final piece in the set of SMPTE ST-2110 standards for professional video over managed IP. Further treatment of the evolution of the video distribution network (VDN) is delegated to Appendix 3.

3) Criteria for selection of CSPs' networks within scope

SDOs' activities and empirical evidence from the dominant application domain of IP video lead to the conclusion that networks within scope are networks that meet the NGN criteria:

- packet – based
- capable of telecommunications at standard levels of QoS
- capable of layered services.

This set of criteria cannot be used as a strict filter, since "legacy access technologies" [134] are still in widespread use. The PSTN – circuit-based and dedicated to voice traffic – is still used and SDH/SONET – circuit-based – is still used in aggregation and in legacy enterprise access. These criteria are intended to guide the analyst in selection of networks that remain relevant as the subject of an energy-use analysis, by virtue of these networks' migration towards the architectural principles of the GII.

4.2.3 Selection of Reference Points – RPI-N or RPI-S?

There is some standardization in meaning and labelling of reference points. ***Reference points for interconnection*** are defined [101] to support the ***interconnection*** of the many actors/players that is at the foundation of the Global Information Infrastructure (GII). What may have appeared visionary at the time (November, 2000), has become the trunk from which all else branches: multiple players in a progressively, increasingly open field. The demand for interconnection is clear and strong.

At the same time, there is sufficient diversity to pose this as one of the difficulties to tackle, i.e., which set of reference points to use. Specifically, RPI-Ns (reference points for interconnection – network; see ITU-T Y.140 [101]) have been chosen to segment the metro-area network. The choice of RPI-Ns over RPI-Ss is guided by several factors:

- The transport is the common denominator across all services.
- The transport is comprised of physical elements which cannot be rapidly re-deployed but carry out their function in a specific place.
- Every service holds a dependency relationship (in the universal modelling language (UML) sense) with the transport.

It therefore remains to determine which criteria guide the selection of the RPI-Ns.

4.2.4 Which criteria guide the selection of RPI-Ns?

Reference points divide functional groups [136]. This separation immediately places the interface points [136] on either side of the reference point, at physical implementations which have different functions. In addition, the interface points, and the physical implementations that hold them, must fit into recognizably different segments. This follows from the definition of the segment in [100]; see especially [100, Para. 6.1.2.9], and the several examples in [137]. Therefore, a useful reference point is one where the interface points lie in different segments. Such a reference point regards not only the relationship between the function of the interfacing groups but, more importantly, it regards the relationship between the function of the segments.

Two levels of function can be distinguished: those in the transport stratum and those in the service stratum. Here, “service” and “transport” are used in accordance with the ITU-T’s general reference model [127]. To classify functions accurately within these strata, it is necessary to use network layers. The reference model identifies (OSI) layers 1, 2 and 3 as possibly within the transport stratum. ETSI’s functional architecture [130] tends to be more specific, referring to “a service layer and an IP-based transport layer.” This inclusion of layers 2 and 3 is problematic. Functions at these layers may employ logical constructs that suit the subsuming (higher-layer-) service’s architecture. As indicated earlier, logical functions may conceal energy users and therefore must be resolved into physical functions. This does not mean that logical functions can be ignored. It means that a reference point suited to the energy analyst must have recognizable interfaces at layer 1. The ITU-T’s guidance on the selection of reference points seems to concur. The GII terminology Recommendation [128] defines a reference point as a point of conjunction of functional groups. It then refers to recommendation I.112 [136] for a definition of the functional group; the latter is identified as possibly “performed by a single equipment”, without any alternative suggested. Such a correspondence simplifies the energy-use analyst's work, as it facilitates allocation of burden. ***This, then, is one criterion: a reference point must correspond to a physical interface***. A direct reference to this consideration is found in ITU-T Recommendation Y.140 [101, p. 8], where it is pointed out that "virtual interfaces" may occur between "non-adjacent network elements". The intermediate network elements cannot be ignored; therefore, efforts to produce a reference model must resolve architectures until all network elements are represented.

Summarizing, concern lies in building a reference implementational model of an NGN, which shows reference points at the location of layer 1 transport – functions (e.g., switching and transmission). A model segmented in this way can be used independently of the services which the network conveys. Furthermore, since RPI-Ns have been standardized at least since Y.140's ratification (2000), their use significantly alleviates the impact on study scope introduced through dependency on the criteria. Provided that a telecommunications network can be segmented according to a model constructed using RPI-Ns, then the depth of individual analysis, as well as cross-comparison between independent analyses, is improved.

4.2.5 Compiling, tabulating and applying RPI-Ns

During proceedings to devise this method, it was observed that the RPI-Ns defined by the ITU-T do not cover the entire range of the telecommunications network. Recommendation Y.140 centralizes the concepts and extant references, but once the due diligence of following through the references is over, then gaps are visible. ITU-T Y.140's scope and purpose regard extant ITU-T recommendations about interconnection's technical dimension. The recommendations referred to are listed in Y.140's supplementary content (notably A.3, B.3, C.4, D.3 and E.2). The span of the reference points only ranges from the terminal equipment (TE) up to interconnection between the access network and the switches that aggregate (upstream) and distribute (downstream) the traffic. Indeed, the ITU-T intentionally provides a framework and an architecture for further development, complementing these with examples in ITU-T Y.120 Annex A [137]. It was therefore necessary to determine whether other complementary standards (possibly, from other SDOs) had laid more groundwork. The more the number of mutually compatible RPI-Ns found, the richer the theoretical foundations of this thesis. By "mutually compatible", it is intended that the RPI-Ns be consistent, coherent and, ideally (if from other SDOs), in acknowledgement of the ITU-T's basal work.

Table VI USEFUL REFERENCE POINTS AND THE SOURCES OF THEIR DEFINITION

Reference point	SDO	Reference
S	ITU-T	I.411
T	ITU-T	I.411
VB	ITU-T	Q.2512
W	Broadband Forum	TR-144

Fortunately, the Broadband Forum (BBF) has taken the approach of extending the ITU-T's work. For example, the Integrated Services Digital Network (ISDN) reference points S and T are used compatibly in TR-144 and TR-145 [133], [134]; furthermore, the W reference point is added [133]. See Table VI . More such RPs are found in later standards from the BBF, and will be shown in Chapter 6.

4.3 Analysis

4.3.1 Stage 1 Segmentation: Edge – Core

In this simple model, the network is divided between an edge and a core. The edge is where a provider's network has connections to an end-user's network. The core is the region of transit between two edges.

This edge-core dichotomy is used extensively in the IETF's treatment of provider-provisioned virtual private networks [20]. Another use of this model is found in the ITU-T's Global Information Infrastructure Recommendation Y.140 [101]. A similar use is seen in the model implicit in RFC 3439 [138], which states the well-known smart-edge, dumb-core networking paradigm. RFC 3439's description of the core of the network is similar in this sense to the packet core of a CSP's network: speed is paramount in the core. Despite the simplicity, there is disagreement. The RFC refers to an edge that is typified by "computers with operating systems, applications, etc.". Thereby, RFC 3439's edge extends beyond intermediate systems and into the communicating end systems. Contrast this with Y.140's limitation of the network's scope. The term user-network interface (UNI) is used to delimit the network, yet Y.140 refers to an edge that is solely within the network's scope.

All edge-core models exclude any direct consideration of the access and aggregation components of a CSP's network. This is not surprising. Abstraction of (detachment from) the grittier details of implementation assists an emphasis on principles which must be designed into telecommunications networks to enable specific fundamental properties. For example, RFC 3439 is intended to guide design for the Internet "backbone". It is concerned with the Simplicity Principle and the delegation of complexity in interworking away from an investment-heavy switching & routing core and towards the switching & routing edge. This simple model is indeed useful; for example, it is an underpinning of IP/MPLS networks. IP/MPLS networks implement policies at the switching and routing edge: path diversity (according to service type) is obtained through decisions taken at the edge, while core nodes are limited to high-speed data plane switching.

Stage 1 segmentation will be used to refer to this modelling approach. The Stage 1 segmentation model ignores the end systems (unlike RFC 3439) and considers the edge to be the boundary of the CSP's network. Fig. 18 illustrates the coarseness of this segmentation. Little structural information is available. Another illustration of application of the Stage 1 segmentation is found in ITU-T Recommendation E.800 [139]. Like RFC 3439, it carries an interest in abstracting the telecommunications network. The Stage 1 model is helpful in such contexts, as it emphasizes properties. In the case of RFC 3439, one such fundamental property is scalability through simplicity. In the case of E.800, it is Quality of Service (QoS).

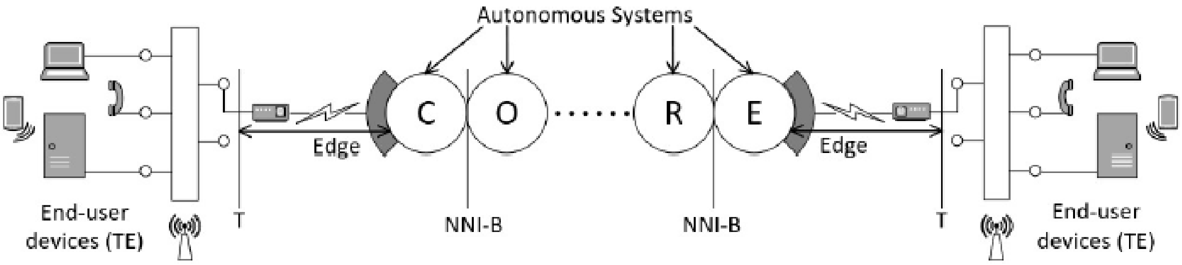


Fig. 18. Stage 1 segmentation of a telecommunications network

4.3.2 Stage 2 Segmentation: Access – Core / Edge – Core

While Recommendation E.800 still manifests the two-stage dichotomy, it introduces and gives meaning to an access segment [139, p. 1]. The access segment appears in this first refinement after Stage 1 segmentation. The edge segment is either replaced by this access segment or separated from it with the resulting downstream segment referred to as the access network. See [16, Fig. 1], [140, Fig. 1], [141, Fig. 1].

This is the access segment which is familiar in the mass market, where references to "cable", "fiber", "ADSL", "4G" and "5G" are easily recognized by the general public. An important distinction to make is whether the work under consideration includes or excludes the segment-terminating device – the CPE – within the access segment. This resolution of the edge into edge and access results in a three-segment model of core, edge and access. This is the Stage 2 segmentation (Fig. 19) model. Stage 2 is an improvement over the resolution of Stage 1 because it reveals implementational detail about the part of the provider's network that physically connects to the end-user's network.

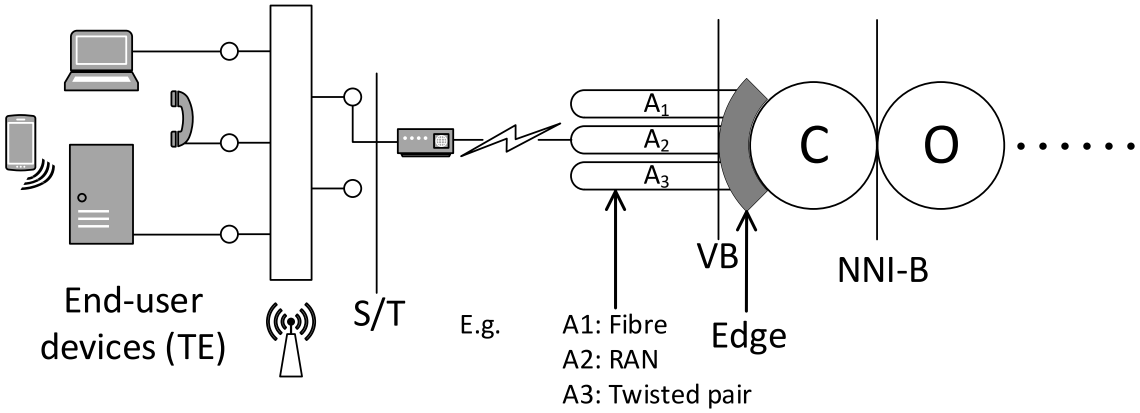


Fig. 19. **Stage 2 segmentation of a telecommunications network**

4.3.3 Stage 3 Segmentation: Access – Metro – Core

1) Variant 1: The Transport Edge

Further resolution is obtained through the functional description of a "metro" segment (e.g., [142, Fig. 1]). In this form, the function of the metro segment is that of aggregation of upstream traffic (and distribution, in the downstream sense) from the access segment. A new meaning for edge is described. It (is that part of the metro segment that) aggregates traffic from the "access network multiplexers". This edge is a thinner segment than that presented in Stage 2 segmentation. Indeed, this "metro" corresponds to what the Stage 2 model presents as the "edge". Furthermore, comparison between works from the same source [140], [142] reveals that the core segment [140] corresponds to the agglomeration of long-haul and undersea segments. The level of sophistication of this form is greater than that of the Stage 2 form. This will now be referred to as **Stage 3 segmentation, Variant 1** (Fig. 20). It consists of the following segments:

1. access
2. edge
3. metro
4. core (long-haul)

This segmentation is also used elsewhere, albeit with different terms. For example, "edge" is variably substituted by "edge aggr." and "metro/edge" [143]. Other common alternative terms are metro-edge or metro-aggregation [143], or simply aggregation [40]. To distinguish these from the metro segment, the latter is then referred to as metro-core [40]. Note that the function of the edge (as part of the metro) in this variant is resolvable only in the transport stratum.

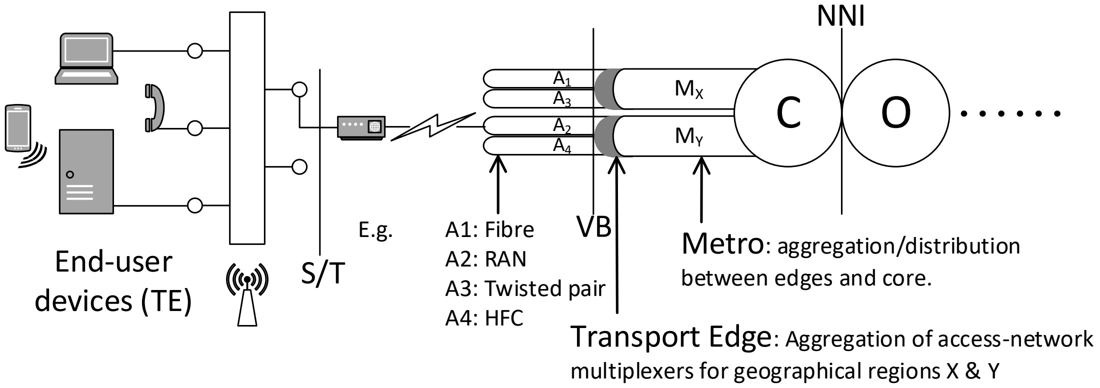


Fig. 20. Stage 3 segmentation of a telecommunications network

In [142], this edge is described using the term "provider edge". This use does not reconcile with defined uses of "provider edge" (PE) (see, for example, IETF [20] and MEF [111]). In [20] and [111], the term PE relates to equipment that implements ***a service interface***. The emphasis is important to distinguish PE equipment from that used to provide transport. This distinction is made in observation of the fundamental separation of service control and provision from the underlying (transport) network of NGNs laid out by ITU-T Recommendation Y.2001 [129]. While the implementations of PE depend on the service, these implementations have a common abstraction that differs from this Stage 3 segmentation (only) in the function and position of the edge. ***Since position is involved, the difference is significant and consequential to the energy analyst, as it impacts the system boundary.*** This abstraction will therefore next be classified as Variant 2 of the Stage 3 segmentation. The discrepancy referred to at the beginning of this paragraph can now be articulated: the use in [142] is irreconcilable because it is not (as defined in [20] and [111]) a ***service interface***. Neither does [142] place a PE router at the boundary of the metro network. According to RFC 4026 [20], which is a systematization of provider-provisioned virtual private network (VPN) terms, the PE router lies at the edge of the service provider network. Note that the PE router may be referred to in other contexts as an Access Router and that the service provider here is the same as that described in ITU-T's (e.g., Y.110, Y.140) service provider domain. This service provider may be a (vertically-integrated) CSP, or it may deliver a telecommunications service as a client of a CSP.

2) Variant 2: The Service Edge

The second variant includes an edge segment that has a service function: hence the term ***service edge*** [144, Fig. 12.1]. In this variant, the edge is the limit of L3 service; it is where the PE routers defined in RFC 4026 [20, Sec. 5.2] are located.

It is important to emphasize that "service" refers to the GII's infrastructural role of "Communication and networking of information" [100, p. 10], not to that of "Generic communications service provision" [100, p. 10]. This latter emphasis is made to assist in the clear separation of network service providers (the yellow bars in Fig. 8) from the underlying network provider (again, see Fig. 8). Without this separation, the concept of the service edge does not obtain sharp relief. A common example of equipment that performs its functions in the service edge is the Broadband Remote Access Server (BRAS). This is where "the last IP aware device between service providers (ASPs and NSPs) and the customer network" in a broadband network, is commonly found [145, p. 7]; indeed, "[t]he BRAS can perform several logical functions [,including that of] … MPLS PE router" [145, p. 7]. Relatedly, reference is made to an "IP edge" in [16, App. S3]. While there is some loose use of terms in this latter source, the graphical aids, notably [16, Fig. S3.8.1], suggests that the Service Edge is intended. The abstraction is illustrated in Fig. 21.

At the service edge, the active forwarding method may change from use of IP header to use of MPLS label, such as when an Internet-route-free-core is operated. This dispensation with BGP, and with IP header lookup inside a single organization's network core has significant implications for external observations of the number of nodes in a route. Both the means (technical) and the interest to conceal core nodes exist. This will be dwelt upon further in the section presenting recommendations.

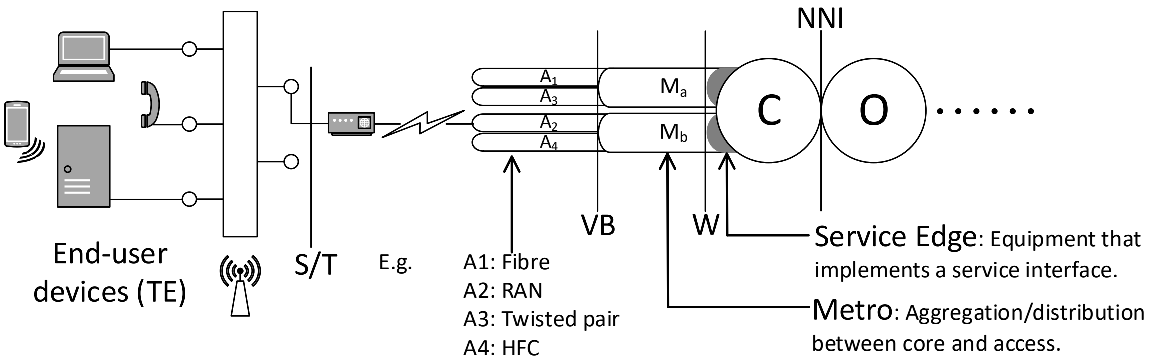


Fig. 21. **Stage 3 segmentation of a telecommunications network, Variant 2 (Service Edge)**

4.3.4 Stage 4 Segmentation: Access – Metro-Aggregation – Service Edge – Metro-Core – Long Haul

Figs. 19 – 21 show a "core" without removing a single layer of its abstraction. This conceals some important implementational detail. Fig. 21 starts to reveal some of it, by identifying the service edge. This sub-section uncovers more detail.

1) Highly dynamic architecture – the metro-core segment

Several local networks typically serve a single metro area. The scope of a local network encompasses more ICT infrastructure than the "set of routers" referred to in the definition of the Autonomous System (AS) quoted in RFC 1771 [146]. It includes elements that are transparent to routing but are not transparent to the energy analyst. For example, transport devices such as transponders, muxponders and reconfigurable optical add/drop multiplexers (ROADMs) – which form the optical network layer (a server layer) – are not needed in representations of AS routing schematics, but they ***are*** needed in accounting for energy use. The optical network will be dealt with more formally in Chapters 5 and 7.

Within the metro area, these aSs interconnect within the system boundaries of the metro-core. This is the part of the core which will now be tackled to complete the reference model of the NGN within the metro area (refer back to sub-section 4.2.2, especially 4.2.2 (3) for modelling scope). The high-level objectives of the metro-core are commercial and technical. Consider three Internet Service Providers (ISPs), say ISP-Local-A, ISP-Local-B and ISP-National-C. Both ISP-Local-A and ISP-Local-B are present with comparable customer bases in the same metropolitan areas but rely on ISP-National-C for traffic with respect to other metros (inter-metro). Through traffic-symmetrical connections known as peering, ISP-Local-A and ISP-Local-B save money that would otherwise be spent on traffic-asymmetrical relationships known as transits with ISP-National-C. Furthermore, peering reduces the latency of packets exchanged between A's and B's customers within the same metro (intra-metro). Control over routing and better bandwidth between the respective local ISPs' customers are two other useful technical advantages of the metro-core. There are several other specific benefits [147] of the metro-core that may be classified under the broad headings of "commercial" and "technical" respectively. Possibly, the pithiest description of this part of the metro-core's scope is that it is the public Internet's market place (as well as that of private internetworking). The metro-core is the segment of the Internet (both public and private) within which connections are made to bypass the use of the long-haul connections [19], [147]. The long-haul ("Tier 1") connections are bypassed because (a) traffic passes directly from source to destination through the metro-core and (b) caches in the metro-core save the repetitive use of the long-haul for the same content. Several other commercial drivers for connectivity in this segment are described elsewhere [19] and the resulting topology is flatter [19]–more mesh-like than tree-like. This growth in connectivity helps to explain why growth of traffic in the metro-core exceeds that in the long haul [46].

It may be helpful to return to dwell on the concept of the segment, in the context of the separation of an apparently contiguous part of the network into two segments. The metro-core and the IP service edge are ***segments***, in observance of ITU-T's definition that "a segment is part of one role, owned and operated by ***one player***, part of one (and only one) service provisioning platform, and part of one domain" (my bold, italicized text). Note that the definition does not preclude segmentation within a service provisioning platform, or within a single player's ownership and operation, or within a domain. Rather, a segment must not cross the gap between service provisioning platform, or that between players, or that between domains. However, in the preceding paragraph, it has been observed that "[w]ithin the metro area, these aSs interconnect within the system boundaries of the metro-core". Therefore, while device equipment boxes can be separated along the metro-core segment's system boundaries, real estate cannot. For example, IX datacentres host several players, interconnecting with one another. Separation of the metro-core from the IP service edge emphasises the functional difference of the two segments; the functional difference is described in the recommendations section. This separation, along with the rest of the model, is shown in Fig. 22.

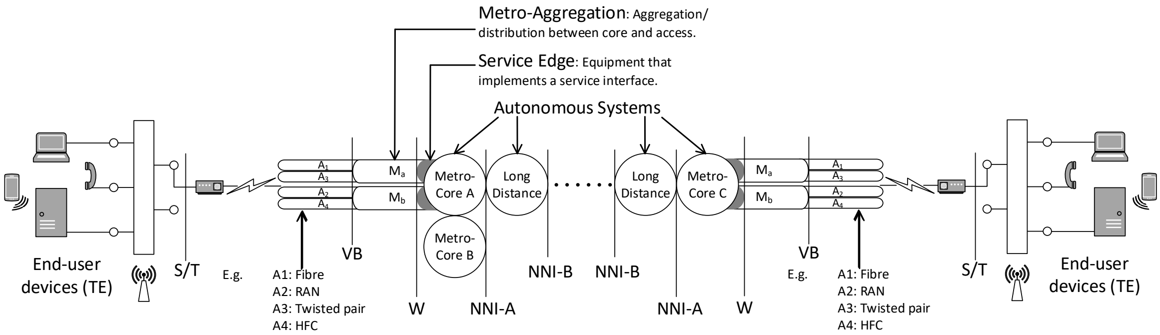


Fig. 22. Stage 4 segmentation of a telecommunications network

2) A stable architecture – the long-haul segment

In the long-haul, whether overland or undersea, the core has the simplest and most stable architecture of the segments. The only technology that meets the bandwidth and latency requirements is optic fibre transmission. Dense-wavelength-division-multiplexing exploits the capacity of the medium to meet current and future demand. Active components inserted in-line with the long-haul cable runs are erbium-doped fibre amplifiers (EDFAs). ROADM nodes terminate the fibre after its runs to connect regions, nations and continents.

3) "Core" – another overloaded term

The term "core" has been used thus far in implementational models presented, and have cited works that include it in their models. Core has at least two classes of meaning with broad implications. The first class [141], [142] takes the term core to be synonymous with "IP core". This use is very similar to that described in the Stage 2 model. Works that use this class of meaning may then elaborate, for example, by referring to the core as comprising other segments, e.g., metro and long-haul [142].

In the second class of meaning, "core" is that part of the telecommunications network that has a control function in addition to a transport function [100, p. 32]. In control, it is the part at the "core" of operations. Operations-categories within this core's scope are aggregation, admission/authentication, switching, billing and service gateways. This class of meaning is evidently service-oriented. This orientation tends to make this meaning of "core" unsuited to inclusion in an implementational model that segments at RPI-Ns (and not at RPI-Ss). This difficulty is implicitly acknowledged in [148, p. 1], through the step of including "control and core nodes" as part of the access network, despite the evident physical separation of the access segment and that which contains the physical equipment comprising the core. As the displacement of circuit-centric operations by packet-centric operations progresses, the two classes of meaning will become indistinguishable.

4) Aggregation – often subsumed

I have shown, in the earlier Stages, that this part of the network is often lumped into another segment. If further evidence is needed, this is explicitly acknowledged in TR-101 [149, p. 13], which, in pursuit of "clarity in this document", defines the aggregation's network span and acknowledges its subsumption into other segments in other Broadband Forum Technical Reports. The position of the metro-aggregation switching segment is now fixed through use of the VB reference point. In the recommendations, I propose that metro-aggregation's share of the burden demands characterization separate from other segments.

4.4 Recommendations

4.4.1 Relevance of Stage 4 to the energy-use analyst

The energy-use analyst's choice of model affects the comparability of results and the capability to detect trends. As has been shown, the decision regards the amount of detail of the physical architecture which is exposed by measurement of energy absorption. A fully detailed model that accounts for consumption by every physical device chassis (box) in the network is likely to prove impractical. Stage 4 segmentation strikes a useful balance between abstraction and implementation for the analyst of energy use.

On one side of the balance, it favours implementation by describing segments that possess a degree of independence in energy efficiency. For example, the energy efficiency of the access segment is technologically independent of that of the metro segment. The technology can be changed without changing the technology used in the metro segment; this is visible in the diversity of access network options available to network providers planning and implementing Fibre-To-The-X (FTTx). By contrast, it can now be observed that Stage 2 segmentation integrates at least the access and metro-aggregation into a single access segment. An analysis of energy-use trends that employs such a model may mask growth in energy use in the metro-core segment by reduction in the access segment. On the other side of the abstraction – concretization balance, the Stage 4 model favours abstraction by segmenting at abstract reference points for interconnection.

Through its resolution, the Stage 4 model favours identification of energy sinks in the telecommunications network. It separates the service edge from the metro-core. With interest in low-latency communication being high as an enabler of novel applications for 5G, this segment is likely to witness high growth in energy use. Furthermore, the metro-aggregation segment is a potential hotspot too, since it may comprise points of presence (PoPs) which can be utilized as Network Function Virtualization Infrastructure PoPs (NFVI-PoPs). Just as the demand for OTT video has led to the growth in metro-core traffic [147], it is reasonable to anticipate that both the service edge's and the metro-aggregation segment’s traffic and energy consumption are at the beginning of an exponential growth curve.

4.4.2 Define Edge as “Service Edge”

“Edge” recurs in the four stages, reflecting its popularity as a term in network architectures. However, overloading necessarily is accompanied by context dependence. While a reader may be expected to infer exact meaning, the reader’s task may be simplified without undue effort if the term were to be part of a recognizable foundation. I recommend that in energy-use analysis, the term “Edge” be reserved for the “Service Edge”. This use is compatible with, albeit stricter than, the use made in the term “Edge Computing”. Compared with the Multi-access Edge Computing (MEC) white paper’s deployment scenarios, the recommended use corresponds to the MEC server at the radio network controller (RNC) deployment site [150].

4.4.3 Divide Core into Metro-Core and Long-Haul-Core and define as “IP Core”

While investigating the “Core”, two uses of the term “Core” were uncovered. Furthermore, a division into a metro- and a long-haul core, was uncovered. As the world progress towards a GII, the “Core” will look more like a “IP core” since the scope of the IP core is subsuming the traditional meaning. “Traditional meaning” refers to the “Core” as that part which has both control and transport functions [100, p. 32]. I therefore recommend replacing “Core” by “Metro-core” and “Long-haul/distance-core” and using the meaning of “IP core” to represent the control and transport functions.

End Service Provider Systems (ESPSs) [151, p. 10] are not included within the metro- and long-haul-core segments. End Service Provider Systems are hosted in datacentre server rooms and interface to the telecommunications network at the core. However, there is significant separation between the roles of ESP and CSP; therefore, ESPSs must be included in a separate segment (by definition of the term “segment”).

4.4.4 Instead of CPE, define the customer-facing endpoint

There is no single understanding of CPE. DOCSIS 3.1’s physical-layer specification standard [152, Fig. 3] presents a network and system architecture that gathers all the customer’s IP-aware equipment under CPE, including personal computers. The cable modem is not included in CPE; indeed, the case of equipment that integrates a cable modem with other functionality is described as a case of colocation within a single device, rather than one where the CM is an item of the CPE class.

The use of RPI-Ns facilitates understanding. In TR-043 [153, p. 6], the Broadband Forum explicitly refers to the DSL modem as the B-NT (broadband network termination). Furthermore, TR-043 [153, Fig. 1] shows the B-NT between U and T reference points. On the other hand, while TR-092 concurs by placing the DSL modem between U and T reference points, a first reading of Annex B’s graphics shows the CPE running IP-level software components that are typical of today’s DSL modems.

The key is the date of this standard. In 2001, these components were commonly run on personal computers, not on the modems. This establishes that the meaning of CPE concurs with that used by CableLabs.

The following example illustrates the ease with which misconceptions about CPE lead to incorrect interpretations of results. An analysis that includes a modem as a pure B-NT would nominally exclude the CPE while adding a highly significant amount of energy consumption, as was claimed in [1, p. 226] and indeed shown in the cross-comparison of numerical results. Without clarification, a reader would see the modem and incorrectly conclude that the CPE is in fact included. This happens because the term CPE has strong connotations with modem. The difference between a modem that acts as B-NT and that which acts as CPE is equivalent to the difference which I.411 makes between an NT-1 and an NT-2. The former faces the customer at the T reference point; the latter at the S reference point. Since S and T reference points are defined [136, pp. 2–3], I recommend using S or T reference points instead of CPE. The use of S and T can be justified as follows.

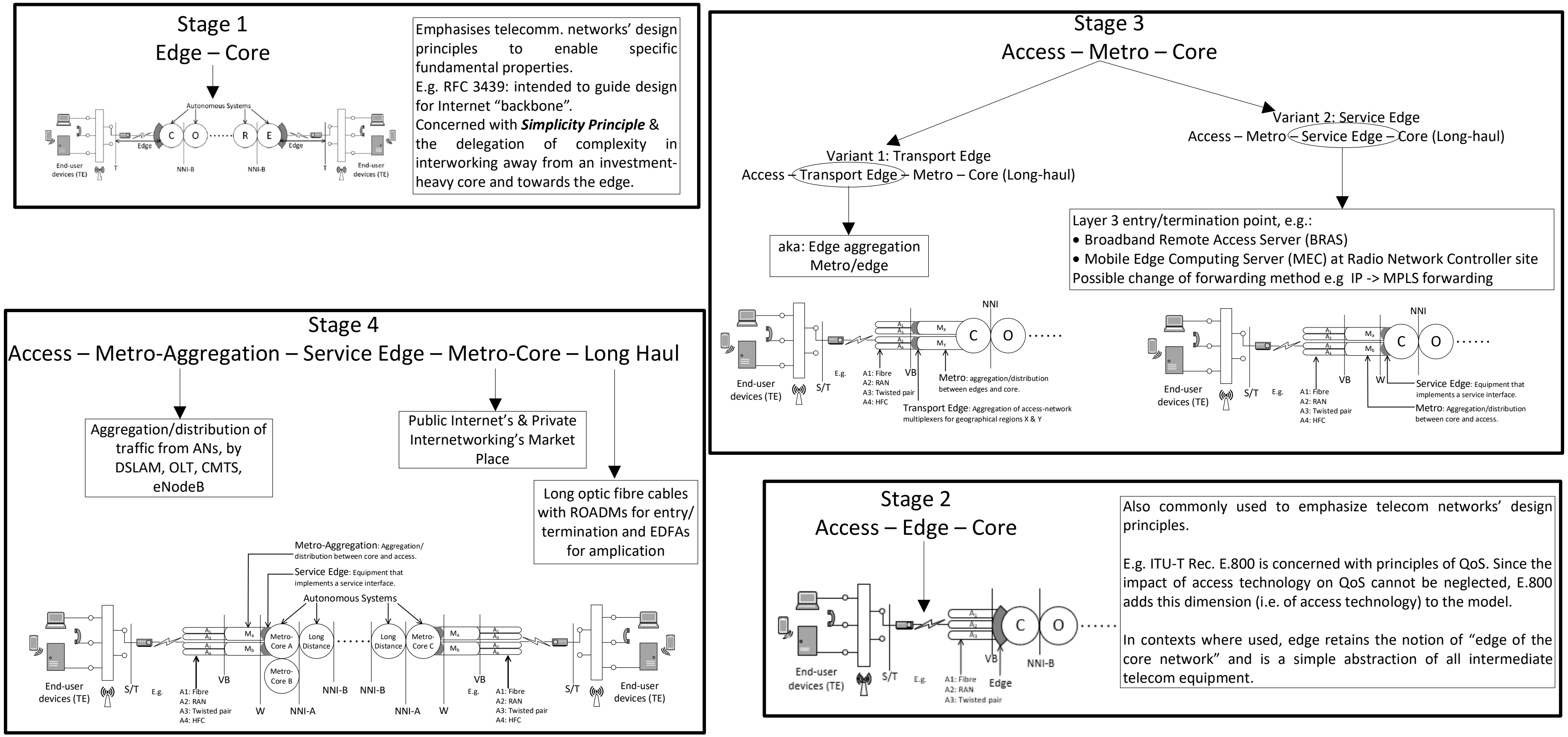
Inclusion up to reference point T ensures that the impact of the telecommunications network is fairly assessed. This is the reference point that is responsible for layer 1 functions and is therefore justifiably attributable to tele-communications. However, in practice, definition of the endpoint is complicated by the fact that enterprise customers may not present it as a physical reference and the tele-communications network's physical property stops at the U reference point. For residential customers, it is common that a residential gateway and modem are co-embedded in the same device. In this case, the S reference point may have to be selected.

4.4.5 Pitfall: External observation of a telecommunications network

An NGN comprises both public and private (managed) parts. By public, reference is being made to traffic being carried on behalf of third parties, i.e., where one of the endpoints is neither the CSP, nor the CSP's customer. The complement is private communication, including IPTV, VPNs and – increasingly – datacentre-interconnection traffic passed over reserved capacity of the long-haul segment [147, pp. 26–27]. This traffic is not accessible to external observers and analysts must ensure that these important types of communication are included in their estimates of energy use.

4.4.6 Fallacy: Use Traceroute to count nodes (hops)

A good example of how logical networking masks energy consumption is found in the use of traceroute. Traceroute depends on the TTL counter in the IP header field, but ASs using MPLS in their IP core may configure their label-switching routers (LSRs) to skip modification of this field. In this case, only the head- and tail-nodes in the path (IP edge) modify the TTL field. The resulting estimate of energy consumption is a severe underestimate.



Stage 1
Edge – Core

Emphasises telecomm. networks' design principles to enable specific fundamental properties.
E.g. RFC 3439: intended to guide design for Internet "backbone".
Concerned with ***Simplicity Principle*** & the delegation of complexity in interworking away from an investment-heavy core and towards the edge.

Stage 2
Access – Edge – Core

Also commonly used to emphasize telecom networks' design principles.

E.g. ITU-T Rec. E.800 is concerned with principles of QoS. Since the impact of access technology on QoS cannot be neglected, E.800 adds this dimension (i.e. of access technology) to the model.

In contexts where used, edge retains the notion of "edge of the core network" and is a simple abstraction of all intermediate telecom equipment.

Stage 3
Access – Metro – Core

Variant 1: Transport Edge
Access – Transport Edge – Metro – Core (Long-haul)

aka: Edge aggregation Metro/edge

Variant 2: Service Edge
Access – Metro – Service Edge – Core (Long-haul)

Layer 3 entry/termination point, e.g.:
- Broadband Remote Access Server (BRAS)
- Mobile Edge Computing Server (MEC) at Radio Network Controller site

Possible change of forwarding method e.g IP -> MPLS forwarding

Stage 4
Access – Metro-Aggregation – Service Edge – Metro-Core – Long Haul

Aggregation/distribution of traffic from ANs, by DSLAM, OLT, CMTS, eNodeB

Public Internet's & Private Internetworking's Market Place

Long optic fibre cables with ROADMs for entry/termination and EDFAs for amplication

Fig. 23. A taxonomy of models observed in literature

4.5 Another recapitulation

In this chapter, it has been seen that the model used in Chapter 3 has the strength of being a good instantiation of a telecommunications network that is ready to undertake the next generation of telecommunication services. On the other hand, it has the weakness of binding to specific technologies and topologies, without significant means of translation to other technical instantiations.

In recognition of that model's weak abstractive properties, this chapter has taken a step towards modelling in a manner that is applicable to a diversity of technical scenarios. This was done through the methodology of implementational modelling. A method was developed therefrom, that took, as its input, a broad range of models observed in the energy literature and has rationalized them with the help of tools published by SDOs prominent in the telecommunications domain. A taxonomy of the models is shown in Fig. 23.

The work of consolidation undertaken in this chapter, is ripe for further development. Three dimensions of development are immediately visible:

1. through application to current – and next – generation technologies;
2. through further resolution of the extent of the metro-area network into segments suited to the use cases that promise to emerge as SDN and Segment Routing are brought to bear, and
3. through inclusion of other structural and infrastructural roles of the GII.

With regard to the third dimension, ITU-T Y.140 claims [101, p. 14] that SP – PTNO[7] interfaces are not standardized. Since the number of infrastructural and structural roles has increased with the advent of edge computing, this condition bodes well for the relevance of this thesis.

---

[7] Here, "PTNO" refers to what is, elsewhere in this work, referred to as the CSP, while the "SP" refers to what is more commonly known as the "OTT" provider. The desired interfaces are therefore those where the provider with some content type to deliver, meets the CSP.

Chapter 5. An implementational model of NGN access networks

Five dimensions have been identified to the challenge at hand. Chapters 3 and 4 mark two steps in the evolution of the effort to meet this challenge. Chapter 3 presented a broad technological and topological instantiation that includes (representative) energy consumers of network elements (NEs) in NGNs. It also identifies the roles involved in provision of telecommunications services. Chapter 4 recognized the limited insight by which an unprivileged energy analyst (one who does not have access to intimate knowledge of network instantiations) is constrained. It therefore distances the reader from instantiations, yet maps out the instantiations using the granular approach of implementational modelling.

The energy analyst is interested in attribution of the burden of energy consumption by telecommunications networks. For energy analysts, ***the physical viewpoint is essential***, since this viewpoint is mandated by the object of study (i.e., energy consumption). If an abstraction even partially conceals the presence of physical entities, then it distorts the accuracy of statistics compiled on the basis of models that use such an abstraction. One such example has been referred to (in an earlier sub-section) as the *traceroute fallacy*, since it implicitly neglects the consumption of devices which are transparent to the time-to-live header field (in both IPv4 and IPv6), such as, for example, the provider-core multiprotocol label switching (MPLS) P-routers. Both layer 2 virtual private network (L2VPN) and layer 3 VPNs (L3VPNs) over MPLS effect this abstraction, which must be resolved for accurate accounting. The means to obtain this resolution are limited; this, too, has been acknowledged in an earlier sub-section. Now, the focus of the physical viewpoint is greatly sharpened by the ***implementational model***, since it “shows which functions are implemented in which equipment” [100, p. 29]. Therefore, it is ideal for the purposes of the energy analyst because:

- the item of equipment represents the limit of granularity which macroscopic analyses are limited to, and
- it facilitates cross-comparison of analyses by a focus on ***functions***, thereby supporting one of the core abstractions of the theoretical framework.

However, rigorous implementational modelling in standards, to date, is weak. Justification follows. Consider Fig. 24, which reproduces a selection of line diagrams from various Broadband Forum (BBF) and ITU-T standards. The scope of comparison here is the **`access network`**[8] (AN), as the ITU-T graphics do not extend beyond that.

---

[8] Courier New font is used to identify terms having standardized definitions and that are rigorously interpreted according to these definitions.

1. A first remark here necessarily concerns the disagreement in terminology between ITU-T G.902 [154, p. 3][9] and BBF TR-101 [149, p. 14], where the ITU-T defines the **`access network`** as extending between the T and V Reference Points (RPs), while the BBF's definition includes the span between the U RP (which may not, or may be exposed, depending on whether it is within an optical network terminal (ONT) or not [155, p. 18]) and the downstream side of the broadband network gateway (BNG).
2. Secondly, marking the subscriber's end of the ONU (optical networking unit) in G.989.1 as UNI (user-network interface) is inconsistent with the ITU-T's own practice (which is itself ambiguous by the ITU-T's own admission – see the results in Section IV).
3. Thirdly, and critically, the line diagrams (i.e., those shown in Fig. 24) define reference configurations that do not account for the energy analyst's essential demand for a physical viewpoint.

Of the five dimensions, it may be fairly claimed that chapters 3 and 4 directly addresses four (at least, to some extent), with the exception being "technological evolution". This chapter moves towards filling this shortfall, by establishing the need:

(a) to update the model with developments in telecommunications network architecture, and
(b) to facilitate application of the updated recommendations through fitting of current- and next-generation architectures onto the updated model.

Rather than concentrating on technologies and techniques, for which reference is made, among others, to [156], [157] in the 5G environment and [158] in MEC, this chapter presents a detailed examination of the architectural paradigms and interfaces that are present in the access network of the metro area, and an attempt at investigation of the impact of MEC and 5G technologies from the point of view of energy consumption. The central contribution of this chapter is that it shows that it is possible to harmonize a wide variety of access network architectures: thereby, it supports cross-operator comparison of energy- and power – consumption data, as it dismantles the energy analyst's central affliction of establishment of comparable boundaries.

The rest of this chapter is organized as follows. In section 5.1, the problem is formulated in detail. The approach is outlined section 5.2 and preliminary results of its application are presented in Section 5.3. Section 5.4 carries an analysis of these results, and section 5.5 concludes with a brief yet broad commentary on the chapter's achievements.

---

[9] I give ***detailed*** references, notably to ***standards***, in order to support verification of my claims, as this work is leading towards a new standard.

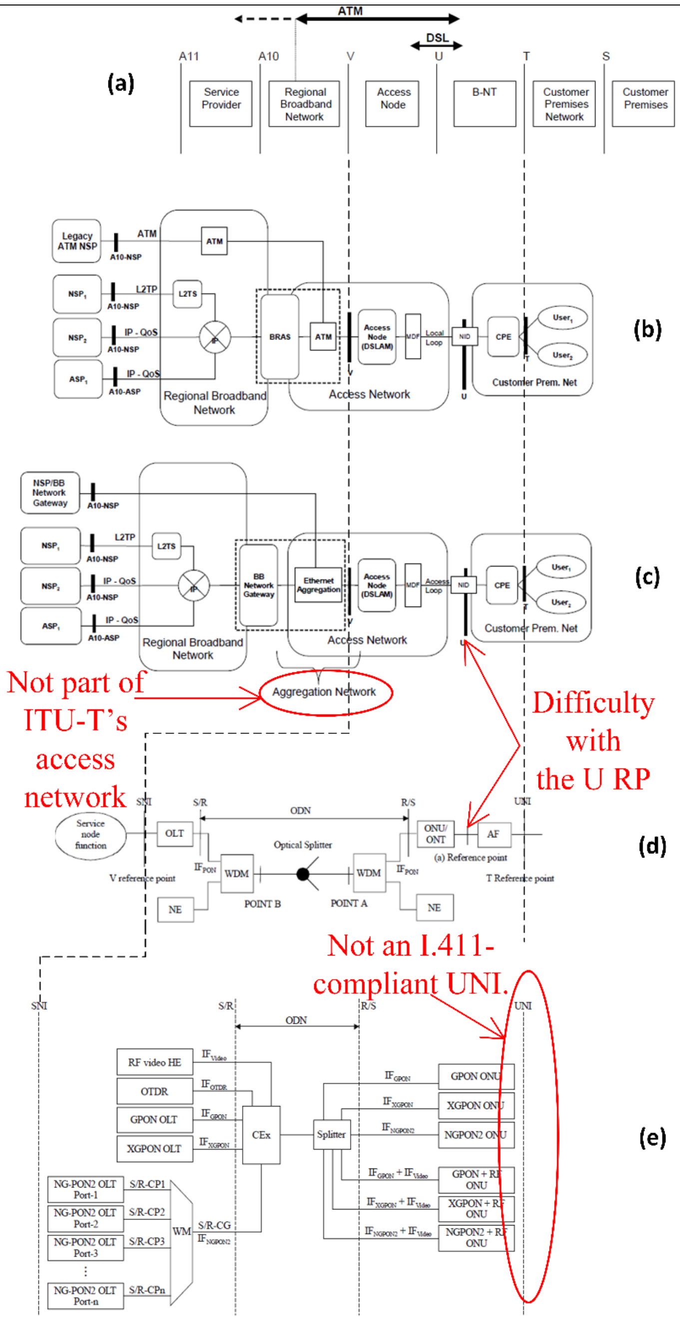


Fig. 24. (Top to bottom)

(a) TR-025 High Level Architectural Reference Model

(b) TR-059 High Level Architectural Reference Model for asynchronous transfer mode (ATM) – based broadband aggregation

(c) TR-101 Issue 2 High Level Architectural Reference Model for Ethernet-based broadband aggregation

(d) ITU-T G.984.1 Functional reference architecture and RPs for GPON

(e) ITU-T G.989.1 Functional reference architecture and RPs for NG-PON2, co-existing with legacy access networks

5.1 Detailed problem formulation

Five aspects (sub-sections 5.1.1 – 5.1.5) will now be revealed of how 5G and MEC:

(a) introduce (as yet un-standardized) physical interfaces, and

(b) add new patterns to the traditional patterns of traffic flow.

These changes require definition of an architectural model as the basis of reporting on energy consumption, yet this model must preserve the physical viewpoint. A review of the theory in context is useful here.

1. The physical viewpoint can be obtained through the recognized means of **`the implementational model`** [159, Sec. 9] with demarcation by (but not limited to) **`reference points for interconnection – network`** (**`RPI-N`**s) [101, Sec. 7].
2. **`Segments in telecommunications networks`** are identified in [159, Sec. 9.3] as consisting of "the customer network segment(s), access network segment(s), core network segment(s), and international network segment(s)" .
3. Each segment is comprised of two **`functional groups`** [160, p. 10]: transport functions and control functions (see, e.g., [161, Para. 5.1] and [159, Fig. 9.2]) (data-plane and user-plane are alternative terms that refer to transport functions).

The segmentation pattern referenced in point #2 is a pervasive theme in standards documents, but in Chapters 3 and 4 it has been shown that the energy analyst requires a resolution of this segmentation into finer reference points. However, there are so many transport variants that it may not be possible to obtain a single, general model. Notwithstanding the need to obtain the abstraction afforded by the **`implementational model`**, it may be necessary to aim for a set of such models that cross-reference the transport network variants to the appropriate model instance.

5.1.1 MEC disrupts current implementational models

The most disruptive of recent developments, in so far as the energy analyst's **`implementational models`** are concerned, originates from multi-access edge computing.

1) A second dimension of energy consumption

The use of MEC in a telecommunications network demands a new investigation of **`RPs`** at interfaces to the **`segments of telecommunications networks`** (defined in [159, Sec. 9.2-9.3]). MEC adds the second, orthogonal dimension of ***computing*** to the transport axis which – until the advent of MEC – has been the only axis along which to align energy consumption.

The **`segment`** is currently characterized as "owned and operated by a single operator", and therefore "interfaces within a segment have a lower priority for standardization than those between segments". While identification of ownership is a secondary concern, the "lower priority for standardization" is an important one. MEC nodes are embedded within operators' (CSPs') **`segments of telecommunications networks`** and they may well be operated by a different organization than the network operator (the CSP). Unless this second dimension is taken into account, the scope for standardization at the {MEC node} – {network segment} boundary, will be overlooked. Lack of standardization at these interfaces hinders efforts to create a universal reporting framework. It leads to aggregation of all consumption in individual **`segments of telecommunications network`**, without the resolution required to locate energy consumers. It is precisely this need for standardized interfaces that is identified in [159, p. 29], where the "interfaces which are important for standardization" include "interfaces between operators and interfaces between equipment from different vendors".

A concrete example of MEC's impact on **`implementational models`** identified to date, can be obtained through inspection of the Stage 4 segmentation model, which was described in Chapter 4. There, **`reference points`** dividing the telecommunications network into access, metro-aggregation, service edge, metro-core and long-haul, were suggested. This effort was undertaken to facilitate interpretation of a telecommunications network from a perspective (that of the energy analyst) that differs, but is complementary to, the perspective that has drawn up **`segments in telecommunications networks`**. The Stage 4 model presents the service edge further upstream from metro-aggregation. This would imply that the first (public -) IP – addressable resource[10] lies upstream from the W **`RPI-N`**. Evidently, this conflicts with the placement of user-IP-addressable (as opposed to operator-IP-addressable) compute and storage resources as far deep as the access segment [162, p. 6] (between VB and S/T **`RPI-N`**s). Affirmation of the changing characterization of the service edge may be found in [163], where the W **`RPI-N`** is reduced to just one of the many possible points where "operational intelligence" (IP addressability included) is distributed over the metro area ("universal metro paradigm"). This observation is reaffirmed in [164] @37:47: the service edge may be located at the cell tower, at the CO/LE or further upstream.

2) Unmapped: where are the resources?

Detailed scenarios of implementation of compute and storage resources are not yet common knowledge [165]; deciding where to locate these resources has been described as "leading to challenges" [166]. The use of granular RPs in such resources' architecture schematics, can help in the extraction of meaning from power consumption statistics.

However, the location of the **`RPs`** is itself unclear. Any segment within the metro area is a candidate for the nomenclative revisionism of the trend to ascribe variants of the "edge" moniker[167]. It is precisely at these edges that some new **`RPs`** must be defined, and/or pre-extant **`RPs`** broadened to incorporate the new interfaces at the edges.

---

[10] This regards IP addressing for services, not for service provision; e.g., GTP-U (**g**eneral packet radio service **t**unnelling **p**rotocol – **u**ser data tunnelling) in 4G and 5G transport typically uses private IP addressing to route tunnelled payloads from external packet data networks to the mobile user equipment.

5.1.2 5G's disaggregated RAN demands more granular implementational models

The 5G System is defined in the 5G system architecture standard [168] as comprising an AN, a core network (CN) and user equipment (UE). The NG `RP` is presented as the junction between the AN entity and the CN, thereby abstracting all intermediate segments that provide ***backhaul***. Since NG is a logical interface [169, p. 8], then this lack of physical detail is not surprising. Indeed, 5GS is ***a logical description of architecture*** [170, p. 76], and a "clear requirement to provide infrastructure connectivity from the Access Points (APs) to the CN, also referred to as *transport network* [sic] connectivity" is observed. This latter assertion extends beyond concern with provision of backhaul alone, as it refers to connectivity from the `Access Point` [161, p. 2]. Such connectivity may include transport across the LLS (lower layer split) (***fronthaul)*** and transport across the HLS (higher layer split) (***midhaul***) as well as backhaul. In this section, I argue for the need to extend characterization of the telecommunications network, in order to correctly attribute energy consumption to energy consumer.

1) Disaggregated RAN deployment scenarios differ in distribution of burden

The problem of architectural rigour in analysis of power consumption is complicated by the diversity of functional splits afforded (for the sake of flexibility in deployment) by the RAN architecture [171, Sec. 11.1]. This flexibility in deployment leads to four different RAN deployment scenarios [172, Sec. 5.4], each with its own unique distribution of burden of energy consumption over the metro area. 4G's baseband unit (BBU) is now disaggregated into 5G's radio unit (RU), distributed unit (DU) and central unit (CU) [172, Fig. 5.2]. The DU and CU may be run as virtualized network functions (vDU, vCU respectively) on "commercial-off-the-shelf" (COTS) server hardware. The most commonly proposed splits between the CU and the DU are Option 7 (the higher-layer split, or HLS) and Option 2 (the lower-layer split, or LLS). Cascaded functional splits "should not be precluded" [171, p. 60]; indeed while a radio unit (RU) is not mentioned in [171], a common cascade is one that includes both Option 7 and Option 2. In this latter case, the RU implements the lower layers and employs evolved common public radio interface (eCPRI) for the transport of packetized data to the DU (or DU/CU). Relatedly, the co-existence of 4G and 5G RANs and transports well into the mid-term future, is another aspect of the same core problem: architectural diversity impinges strongly on the distribution of energy consumption across the metro area.

This flexibility alters the distribution of energy consumption across the metro area and therefore impinges directly on the suitability of the F1 RP (between CU and DU) for inclusion in the implementational model. On the one hand, in the HLS – only case, the vCU may run in the nearest central office (CO) / local exchange (LE) / BBU hotel, with the vDU likely running on COTS at the cell tower/mast site. In this case, the F1 `RP's` scope of abstraction terminates at one end at the V `RP` (demarcation between access and metro-aggregation) in the CO/LE. On the other hand, in the LLS – only case, the vDU is moved out of the cell site and into the CO/LE. This places the F1 `RP` ***within*** the

CO/LE and it no longer coincides with the V RP (V **`RP`** is defined in [133]). In this case, it is possible that the F1 **`RP`** may not coincide with any physical **`RP`** and may be entirely subsumed within a compute host, much as the 5G Core's (5GC's) service-based interfaces (SBIs) may be.

2) The disaggregated RAN demands new RPI-Ns

`RPI-N`s have been indicated [101] as the means to define points of interconnection between ***different organizations***. However, the disaggregation of the Next Generation nodeB (gNB), as well as Open RAN's emphasis on multi-player connectivity, lead to the perception of **`RPI-N`**s ***within*** the **`segments in telecommunication networks`** demarcated by the traditional location of **`RPI-N`**s in the metro area, i.e., the AN interface to aggregation and the IP service edge to aggregation (V and W respectively, see [133]). Although centralized RAN (C-RAN) may appear to present a similar problem in LTE, the Common Public Radio Interface's (CPRI) stringent QoS demands (particularly on latency) have not favoured multi-organization provision or operation. Therefore, prior to 5GS's disaggregation and Open RAN's openness, there had been little to no scope to seek **`RPI-N`**s at interfaces between components of the RAN. In the new RAN context, a scope can indeed be found, and there seem to be no evident **`RPI-N`**s that can fill the role.

5.1.3 Transport is inadequately standardized from the energy analyst's perspective

The **`metro transport network`** is the term recently adopted by the ITU-T[173] to identify a new connection-oriented layer of transport infrastructure, based on FlexEthernet, that can be used in 5G. However, the progress that has been registered to date does not adequately meet the energy analyst's requirement. This sub-section points out the architectural diversity which the energy analyst must reconcile in order to obtain an accurate representation of the physical viewpoint.

1) Segments, xHaul or Domains?

Some justification of this need (to match the perspectives represented by the architectural diversity) comes from [174, Sec. 6]. This section is entirely dedicated to the problem of mapping crosshaul parts (fronthaul, midhaul and backhaul) onto **`transport network domains`** ("metro access, metro aggregation, metro core, and backbone domains"). This departure from the ITU-T's two-segment pattern (i.e., access and core) is needed as there are points of interest and interfaces within the classical **`access segment`**, and these create scope for identification of **`RPI-N`**s to facilitate unequivocal analysis. At least, the energy analyst needs to complement the two-segment pattern with an understanding of underlying **`layer networks`** [175, Para. 6.1.1] [161, Para. 5.2.1.1]. These complementary descriptors are considered next.

2) Transport is recursive

Recent efforts have standardized architecture [173] and interfaces [176] of the path and section layers of the **`metro transport network`**, and emphasized that these layers are non-recursive.

Two cases (passive optical network – PON – fronthaul and optical transport network – OTN – respectively) are considered in some detail in [177], [178] to fulfil the role of the optical media layer in part or all of crosshaul.

However, the ITU-T describes a generic functional architecture of transport networks, and explicitly identifies ***recursiveness*** in the transport network [161, p. 5]. In [111, pp. 16–17], the Metro Ethernet Forum (MEF) elaborates on this by referring to the "dual role" which **`layer networks`** like MPLS play in the Carrier Ethernet stack, acting both (possibly within the same stack instance) in the application service layer and in the transport layer. Such recursion is ***abstractive***. The recursion follows the generic rule of a **`path layer network`** [161, p. 3] that is the client in a **`client/server relationship`** [161, p. 2] with a server layer below it, recursively, until the **`transmission media layer network`** [161, p. 4] is reached. Therefore, consideration of **`transport entities`** [161, p. 4] and **`transport processing functions`** [161, p. 4] of a specific **`path layer network`**, without consideration of the abstracted **`layer networks`**, will underestimate energy consumption.

5.1.4 Agile routing for slice support means unpredictable flows

To date, traffic in the metro area is predominantly logically hubbed, traversing from access nodes (perhaps on an optical ring in **`transmission media layer network`** topology) to aggregator nodes; the aggregators are themselves logically hubbed, on another optical ring, to metro-core nodes, en route to data centres that host the target service. This, coupled with a utilitarian emphasis which depicts an application-agnostic[11] communications service, has led to the nickname of the "dumb pipe".

The status quo is changing. With the increased scope for traffic engineering to support network slicing, transport is becoming "smart"; e.g., capacity is allocated during transport service planning, to enable provision of services with specific quality. With 5G, planning of services is becoming ***agile***, through the use of automated transport network re-configuration. Therefore, scope of study includes the routing technology that takes an application's ***intent*** as input and, as output, configures the transport to meet the application's quality-of-service (QoS) demands.

5.1.5 The implementational model's dependence on technology

Concern here lies with the breadth of diversity that faces an attempt to generalize telecommunications networks through the means of the **`implementational model`**. This problem

---

[11] In fairness, this perception is only true for best-effort services that oversubscribe pipe capacity and rely on various predictors such as statistics (e.g. Mth percentile of N-minute utilization samples exceeds X% of link capacity over a Y-day period), switch output queue drops and demand growth rate [179], for capacity planning.

spans a range that is too broad to treat meaningfully yet concisely. Treatment is limited to two examples, to justify the significance of this aspect of the challenge

1) Topology's dependence on technology

XR[12] optics enable a radical departure from current topologies through the provision of long-reach segments all the way from the residential and commercial **`user-network interface`** (**`UNI`**) [128, Para. 62] , to the network core. XR optics facilitate fronthaul ultra-low latency, which TDM-PON implementations may be unable to provide due to their inherent, lower limit on latency(see, for example, [180, N. @70:18]). While this technology is still nascent at the time of writing, it strongly impacts topological deployments. Notably, XR optics may hold the key to collapsing metro-core and metro-aggregation into a two-stage point-to-multipoint topology [181].

2) IP over DWDM

In [182], Arelion's (ex-Telia Carrier) representative describes adoption of 400ZR pluggable transceivers directly into router chassis, for a metro area network. This collapses the transport stack of **`layer networks`** to a minimal, cost-effective means of transporting IP traffic within the metro area. The contrast drawn here is between the thereby-enabled hop-by-hop routed architecture and competing alternatives, such as agile (i.e., wavelength-switching) ROADM-based (reconfigurable optical add-drop multiplexers) networks and OTN switched connections.

5.2 Approach

The approach consists of an application of the modelling framework of the Implementational Model (see sub-section 5.2.1, below), which prescribes the use of specific topological artifacts (see sub-section 5.2.2), as obtained from relevant standards developed by stakeholder – SDOs (standards development organizations), under the overarching guidance of the five aspects (sub-section 5.1.1 – 5.1.5).

5.2.1 The modelling framework: the Implementational Model

The **`implementational model`** is introduced as an object of standardization in ITU-T Y.110 [159], where the need to balance functional representation of the Global Information Infrastructure (GII) with physical representations, is introduced. Furthermore, in [159, p. 1], ITU-T Y.120 [183] is referred to for a framework of a method for development of an **`implementational model`**. My approach is rooted strongly in this framework (see, in particular [183, pp. 1–2]). Notably, item (b) in Y.120's framework requires "*identification of the set of standards that could be applied at each key interface point*". While consensus is sought within an SDO in the process of development of

---

[12] Variable bit-rate transceivers – see https://www.infinera.com/wp-content/uploads/XR_Optics_FAQ.pdf

a standard, such consensus is confined to the collaborators within the SDO. Where overlapping scope exists across SDOs, it is necessary to cross-correlate the diverse standards, thereby attempting cross-SDO consensus.

5.2.2 Modelling artifacts

1) Partitioning, reference points, RPI-N and RPI-S

An essential development I bring to the ITU-T Y.120 framework is to depart from `partitioning` [161, Sec. 5.3.1.1] of the diverse architectural variants of the `transmission media layer network` (the topological variants of the optical, radio and copper media). The `partitioning` must be guided by the five aspects (Section II) and result in `reference points` that describe the deployment of the metro-area network in terms of physical interfaces.

As the **`implementational model`** does not demand adjacency in the interfaces it shows, use of the **`RP`** is insufficient to guarantee accurate accounting. Here, ITU-T Y.140 [101] is useful, as it defines the concept of the **`RPI-N`**, and distinguishes it from the **`reference point for interconnection – service (RPI-S)`**. The **`RPI-N`** is an interface at a physical adjacency, but the **`RPI-S`** is not bound by this physical constraint. Indeed, the protocols that regulate communication between the **`element`**s [128, Para. 30] on either side of the **`RPI-S`**, may be carried over several **`RPI-N`**s that are intermediate to the two **`element`**s.

An approach based on the key distinction between **`RPI-N`**s and **`RPI-S`**s can be perceived. **`RPI-N`**s are identified at the lowest **`layer network`** – the **`transmission media layer network`** – in order to ensure that all energy consumers are captured. This assurance is obtained from this **`layer network`**'s presence at every network node; without this **`layer network`**, an **`element`** in a higher **`layer network`** (i.e., the **`path layer network`**) cannot communicate with a peer at another node. The **`RPI-N`**s thus serve the dual purpose of capturing the energy consumers and locating the **`reference points`** that frame the layer's topology. However, at any higher **`path layer network`**, the **`RPI-N`** does not exist. Here, the **`RPI-S`** construct fills the role of demarcating of the interfaces between **`element`**s and thus serves the same purpose, i.e., capturing the energy consumers. The process iterates through all **`layer networks`** until all consumers within the service's scope in the metro area, are captured. A recapitulative name for this process could be "serial recomposition of network services"; its objective is that of obtaining an implementational model, populated with reference points that capture all energy consumers. A good example of scope for application of this process would be a Metro Ethernet service (private line, virtual private line, private LAN, virtual private LAN, private tree and virtual private tree).

2) Complementary referential constructs: IrDI, IaDI

One important observation remains to be made. It may not be possible to obtain known **`RPI-N`** and/or **`RPI-S`** constructs in the layer networks, or they may not exist at the granularity required to demarcate energy consumers. Fortunately, complementary **`reference points`** suited to the challenge at hand do exist, at least as generic alternatives, that fit this purpose.

1. To move vertically, across **`layer networks`**, there exist **`access points`** [161, Para. 3.2] that represent the handoff of adapted client layer [161, Para. 3.10], to the **`trail termination source`** [161, Para. 3.43] of the server layer (and the opposite direction, too).
2. To move horizontally, along a **`layer network`**, there are **`inter-domain interfaces`** (**`IrDIs`** – see, e.g., [184, Para. 1], [184, Para. 3.2.1]) and `intra-domain interfaces` (**`IaDIs`**, **`IrDIs`** – see, e.g., [184, Para. 1], [184, Para. 3.2.1]).

3) Between the RPs: topological components

Next, to proceed from reference points to implementational model, **`topological components`** [161, Para. 5.2.1] are particularly useful, as they are obtained "in terms of topological relationships between sets of like reference points" [161, Para. 5.2.1]. Through (a) the use of each **`layer network`**'s reference points, and (b) working through the **`layer networks`** from the bottom up, it is reasonable to expect that all energy consumers are captured, along the transport axis of the telecommunications network in the metro area. Finally, en route from **`UNI`** to metro-core, it is again reasonable to expect the need for (new) **`RPI-N`**s with MEC nodes. Summarizing: an **`implementational model`** can be constructed through abstraction of technological implementations by use of **`topological components`** specific to the **`layer networks`**.

5.2.3 Categorization

In the process of construction of identifiable **`implementational models`** of a **`layer network`**, it should be possible to abstract some differences and obtain categories. These **`layer-network-`**categories can then be combined with categories in the other layers to form bonded verticals through the categories. These bonded verticals will form implementational models suited to the energy analyst's interpretive lens.

5.3 Results: A Unified Reference Configuration at the Subscriber's end

The approach results in a systematic restructuring of diverse implementations of the subscriber's end of the telecommunications network in the metro area, into ***a unified reference configuration at the subscriber's end***. To recapitulate, I reiterate that ITU-T Y.120 recommends *"(a) identification of points that form key interconnection interfaces, access interfaces or appliance* *interfaces in a configuration involving a set of providers of services, networks and appliances; (b) identification of the set of standards that could be applied at each key interface points".*

Fig. 25 and Fig. 26 are the result of execution of the processes that derive from an interpretation of these two recommendations (acronyms expanded below Fig. 26). In my source schematics, the two diagrams are vertically aligned in one continuous layout. Here, to improve readability, the (partial) schematic has been divided along the length and the two parts laid out side by side. The RPs identified (both standardized – shown in bold – and non-standardized), shown at the top of the diagrams, are described next, followed by a tabular summary (Table VII) of the sources. The models are not intended to be an exhaustive reference configuration but they ***are*** intended to facilitate simple extrapolation to match any other possibility of **`access network`** in the next-generation network.

5.3.1 Reference Points

1) **S** RP

This is defined in ITU-T I.411 and affirmed in BBF TR-025, as well as MEF 4 [111]. By "affirmed", I mean that the use made in TR-025 [185] and MEF 4 is recognizably the same as that established by I.411. I.411 and MEF 4 identify the S RP as the point where end-user / terminal equipment interfaces with a private customer network / local area network. End user equipment lies downstream of this RP.

2) **T** RP / **CMCI**

This is defined in ITU-T I.411 and affirmed in BBF TR-145 [134], as well as MEF 4. CableLabs' specification of the modular headend architecture includes a cable modem to CPE interface (CMCI) [186, Fig. 5.3] that coincides with the T RP.

This RP might be referred to as the UNI (e.g. [187, Fig. 2] and [111, Fig. 1]), but Y.120's observation on the UNI's ambiguity (not to mention my tacit agreement on its liberal use as a term) guides us to avoid including UNI in the reference configuration. An incomplete understanding of the T RP may lead to incorrect attribution of the burden of energy consumption between the subscriber and the network or service provider. If, the subscriber uses xDSL access, the T RP may be externally inaccessible and embedded within the integrated xDSL + RG (residential gateway) device. I comment on both these issues in Section V (analysis).

3) **U** RP

This is described in TR-043 with affirmation in TR-101 Issue 2. TR-043 acknowledges that use corresponds to ITU-T practice. However, I.411 explicitly declines to standardize this reference point, with the observation that *"there is no reference point assigned to the transmission line, since an ISDN user-network interface is not envisaged at this location."* Despite the lack of a primary definition (since the underlying reference does not seem to exist), uses made in the BBF documents and popular literature (e.g. [188, p. 321]) are reconcilable.

An incomplete understanding of the U RP may lead to incorrect attribution of the burden of energy consumption between the subscriber and the network or service provider. If the subscriber uses PON access, the U RP may be externally inaccessible and embedded within the integrated ONU + RG device. See section V-B for further analysis.

4) **PAI** / DP / **R/S** / Tap

This point co-locates various references to the network. The premises attachment interface (PAI) is defined in ITU-T Y.120. Its location upstream of the NT (network termination) device in [183, Fig. 5] assists generalization of the PAI's location. The terms "DP" (distribution point – see, for e.g., [189]) and "tap" (see, for e.g., [190, Sec. 6.4]) are used by network personnel to refer to the PAI with more technically-specific meaning than the general "PAI". Since the R/S RP is just before the ONU/ONT (in the downstream sense), then R/S coincides with the PAI. Note that between the PAI and the U RP, xDSL access (excluding G.fast) has no active devices.

5) **DI**/SAI/PCP

This point relates to the "serving area interface", and "primary connection point" (PCP) [189] which I have not found in standards but is well known in technical vernacular. The drop-distribution interface (DI), defined in Y.120, matches well with the common understanding of the serving area interface (SAI) and the PCP (as the point where local loops are cross-connected to feeder cables). Given the universality of existence of this type of point across all wired access networks (see Figs. 2, 3), it is useful to include this in the universal network schematic. Since this point often includes powered equipment, it is reasonable to expect that this RP will indicate a location hosting compute and storage equipment.

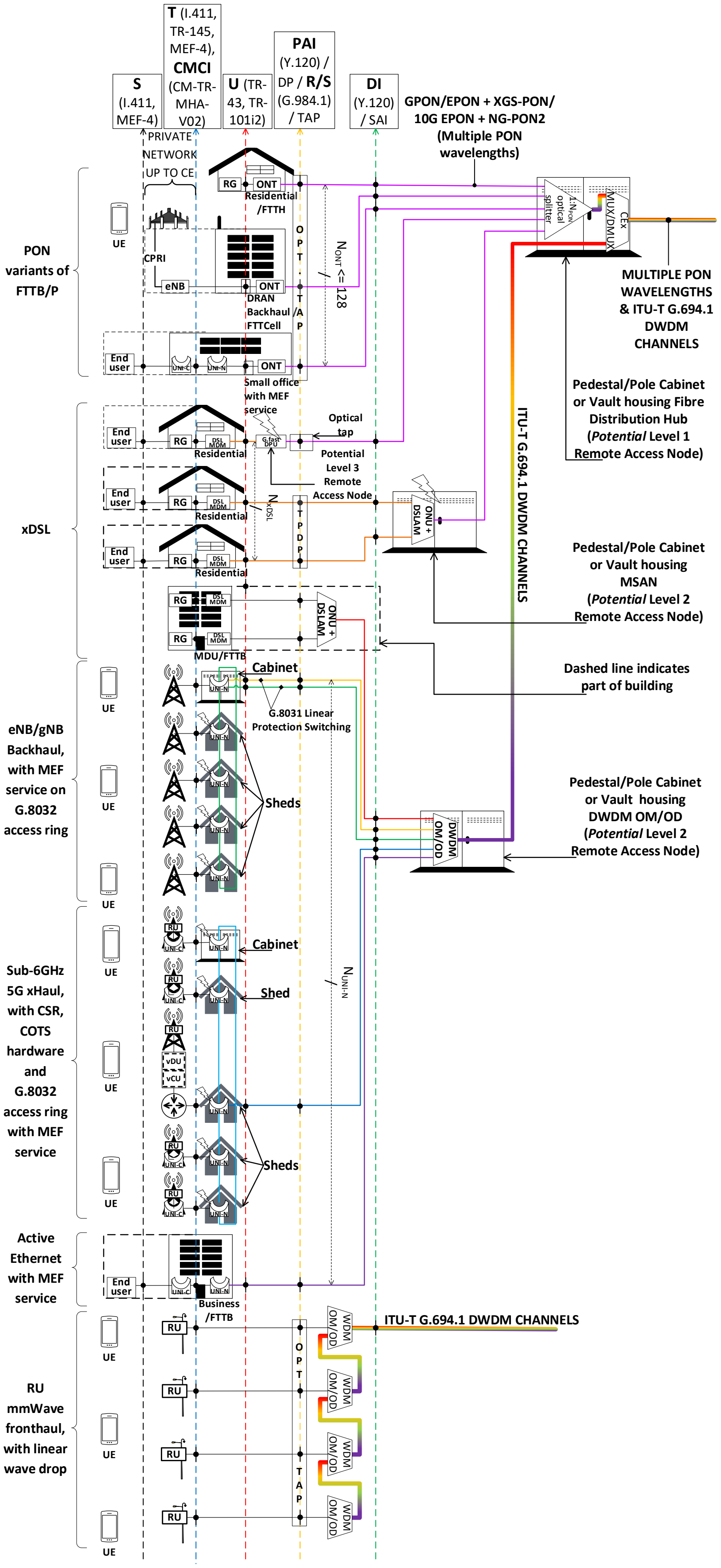


Fig. 25. Unified reference configuration of various access technologies at subscriber's end (part 1)

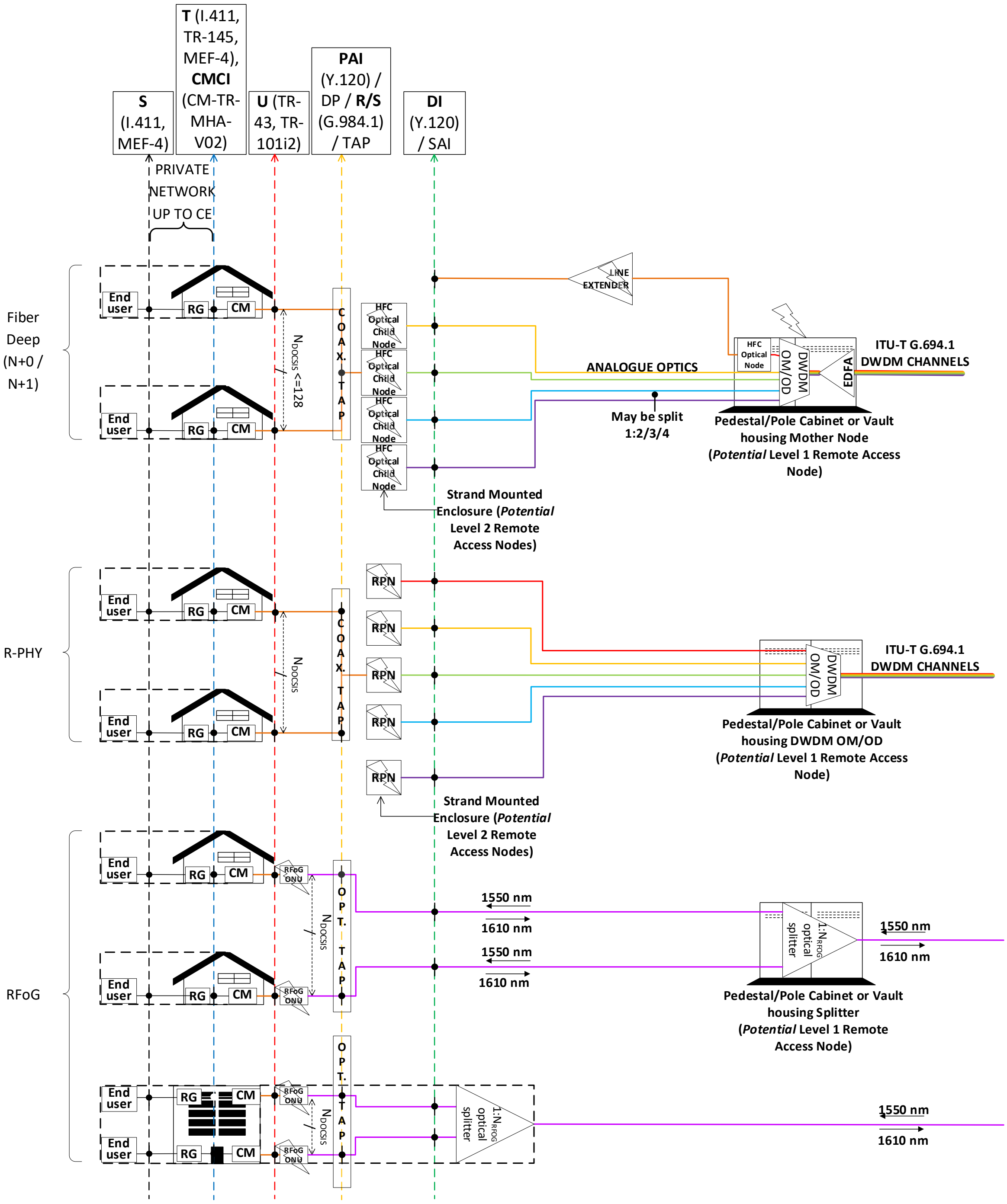


Fig. 26. Unified reference configuration of various access technologies at subscriber's end (part 2)

CM: cable modem; CSR: cell site router; DSLAM: digital subscriber loop access multiplexer; DU: distributed unit; HFC: hybrid fiber-coaxial; MDU: multi-dwelling unit; MSAN: multi-service access node; RU: remote unit; OLT: optical line terminal; OM/OD: optical multiplexer/demultiplexer; ONU/T: optical network unit/ terminal; RFoG: radio frequency over glass; RG: residential gateway; RPN: remote-phy node; RPS: remote-phy shelf; UNI-C/N: user-network interface customer/network

5.3.2 Summary

Table VII RECONCILIATION OF REFERENCE POINTS

RP	Standard	Remarks
S	ITU-T I.411, BBF TR-025, MEF 4	**Original source is I.411.** Use made in TR-025 refers to “ITU-T practice”. Use made in MEF 4 is recognizably the same (refers to I.324, which refers to I.411)
T	ITU-T I.411, BBF TR-145, MEF 4	**Original source is I.411.** In CM-TR-MHA-V02-081209, CableLabs describes a cable-modem-to-CPE-interface and labels it as CMCI. CMCI coincides with the T RP. MEF 4 aligns the T RP with the UNI.
U	BBF TR-043, BBF TR-101i2	Earliest use is made in TR-043. Refers to consistence with “ITU-T practice”, but no such definition has been found in ITU-T standards. I.411 explicitly decines to acknowledge this as an RP.
R/S / PAI / DP / tap	ITU-T G.984.1 / ITU-T Y.120 / BT Openreach WLR3 reference pack / ANSI/SCTE 153 2021	G.984.1 (R/S) regards the optical distribution network; ANSI/SCTE 153 2021 regards HFC; DP mostly regards copper twisted pair. Y.120 is technology agnostic.
DI / SAI / PCP	ITU-T Y.120 / vernacular / / BT Openreach WLR3 reference pack	Y.120 is technology agnostic while PCP mostly regards copper twisted pair.

5.4 Analysis

5.4.1 The UNI – “ambiguous”

I first contrast the functional reference architecture shown in G.989.1 [191, Fig. 5.1] with the result of application of the approach.

G.989.1 [191, Fig. 5.1] (reproduced in Fig. 24(e)) includes the “ambiguous” [183, Para. 8.1] user-network interface (UNI). Fig. 24(e) shows how G.989.1 [191, Fig. 5.1] converges all PON services onto a single implementation. The UNI is shown directly downstream of the ONU. This is problematic because standard G.984.1 defines an **`adaptation function`**: “*additional equipment and/or function to change an ONT/ONU subscriber-side interface into the UNI. Functions of AF depend on the ONT/ONU subscriber-side interfaces and UNI interface*.” Indeed, G.984.1 [187, Fig. 2] presents a **`reference configuration`** [160, Para. 2.4.421] that defines an ephemeral RP, referring to “(a) *(sic)* Reference Point”, and states “[i]f AF is included in the ONU, this point is not necessary.” This latter condition is essential, as it differentiates between (1) the functions of terminating the optical access network (OAN) and (2) the adaptation function.

In [155], the term “adaptation function” is used differently (see, e.g., [155, pp. 17–18]) and overlaps with the scope of the NT1 functional group (see ITU-T I.411 [136]). However, in [155, Figs 2–3], some reconciliation with [136] is obtained, since the meaning of the T RP as established in [136] is recovered. The T RP demarcates the separation between the NT1 and NT2 functional groups. Simply put: the T RP is the nexus of the customer’s local switch and the provider’s NT1 functional group. This diversity is collated in Fig. 27, where the complexity is resolved through alignment of reference points. Through the medium of Fig. 27, it is shown that:

- that use of the UNI in G.989.1 Fig. 5-1 is at best problematic and at worst incorrect;
- the diversity of terms used to refer to the same functional group (e.g., NT1, AF, RG);
- the difficulty in locating the U RP; indeed, this RP may not be externally accessible.

5.4.2 The energy analyst must understand subsumed RPs

1) U RP

The U RP may be subsumed within (internal to) equipment located within the customer's premises. See Fig. 4, and contrast TR-156 Fig. 2 with TR-156 Fig. 3. This subsumption is also hinted at in TR-101 (Issue 2) Fig. 3, where the U RP bisects the network interface device (NID).

2) T RP

The T RP also may be subsumed within equipment located within the customer's premises. In [134, Para. 4.2.1] (TR-145), it is observed that, for broadband access services, the T RP may be "between the RG and other CPE in the customer location or between a B-NT and an RG". Neither of these cases is helpful. As regards location between the B-NT and the RG: current practice consolidates as many functions as possible in a single item of equipment, and it is likely that both the B-NT and the RG are such a single item. As regards location between the RG and other CPE: a description consistent with practice predating this claim, shows this to be the S RP – not the T RP. The S RP is the point where the end user equipment interfaces. Indeed, the RG interfaces with end user equipment, either directly (through an embedded Ethernet bridge or an embedded WiFi bridge), or indirectly, through a private customer network. Notably, the S RP is absent from TR-145.

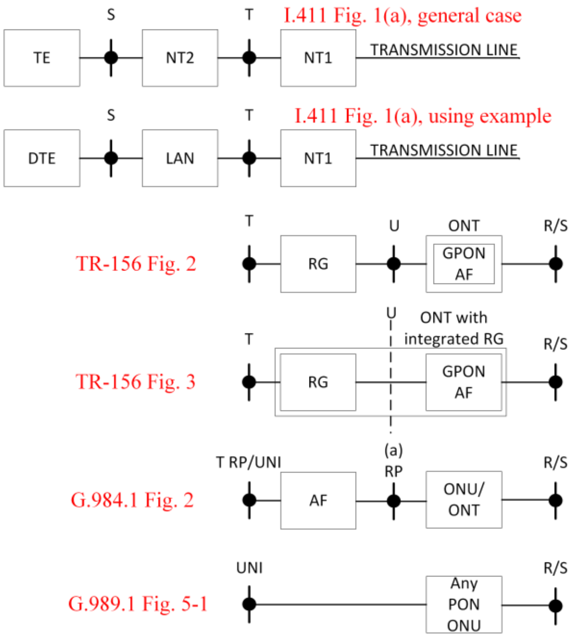


Fig. 27. **Collation of reference configurations that elaborate on the telecommunications network in the vicinity of the subscriber (AF: adaptation function, RG: residential gateway, NT: network termination, TE: terminal equipment, DTE: data terminal equipment)**

5.4.3 TR-145: conflictual descriptions of U1 and T

TR-145 introduces the U1 RP and locates it at the same point where the T RP lies. Indeed, part of the descriptive text for U1 is equivalent to that for T. For U1: "e.g. [b]etween xDSL … and RG function for consumer services. For T: "between a B-NT and an RG." Without delving into an evaluation of whether the introduction of this RP is justified (as opposed to use of existing RPs), the inaccuracy inherent in the above description of the U1 RP must be accounted for should the energy analyst choose to refer to it in an analysis.

5.4.4 Access Nodes are prime candidates for MEC nodes – and MSOs have an advantage

In sub-section 5.1.1, it was shown that MEC adds a dimension to energy consumption that is orthogonal to the transport axis. The MEC nodes need space, cooling, access to the network and power. This sub-section focuses on network and power, touching upon space in the process

1) Addressing "unmapped: where are the resources?"

The model (Fig. 25 and Fig. 26) shows a number of ***potential*** access nodes. Access nodes are defined in [149, Sec. 2.3] (TR-101 Issue 2) and requirements established in [149, Ch. 3]. A fundamental requisite is that it "must have an Ethernet uplink providing connectivity to the aggregation network". Thus, for a site's potential to be realized, it must have an Ethernet uplink. Two broad cases are considered, following which a recommendation is made.

a) Case A: passive distribution

A cabinet (on pedestal or pole) or a vault housing a DWDM optical multiplexer/demultiplexer (OM/OD) may have the required space to host hardened compute and storage equipment; however, the OM/OD is a passive device. Therefore, there is no inherent network facility in this site. Moreover, ***there is no power supply at this site***. The erstwhile advantage of obtaining distribution without the use of power, is now reversed into a disadvantage. The site may, of course, be provided with a power supply and an Ethernet uplink device may be installed. The Ethernet uplink device may, for example, terminate all the wavelength cables issuing from the DWDM OM/OD, and use coloured pluggable transceivers in the downstream ports.

This case also includes (G-, XG(S)-, NG-) PON distribution hubs. These hubs host passive power splitters; inherently, they require neither electrical power, nor a port that frames subscriber data in Ethernet frames for transmission upstream. Both the provision of power and the re-design of the network end of the hub to send and receive Ethernet frames, are significant in financial and technical senses of the word.

b) Case B: active distribution

A similar physical location housing an MSAN may very well aggregate its traffic over an XG(S)-PON. The XG(S)-PON "must provide access to carrier-grade metro Ethernet services" [192, Sec. 7.6]. Such a site does, therefore, satisfy the access node's criterion for Ethernet. This case includes both FTTN (fiber-to-the-node) as well as FTTB (fiber-to-the-building/MDU/MTU (multi-dwelling unit/multi-tenant unit)), where the subscriber-end of the optical distribution network (ODN) is at a cabinet (FTTN) or service room (FTTB), but distribution further downstream is over copper media (e.g., xDSL, or Ethernet PHYs adapted to copper media). In this case, space, power and an Ethernet network are all provided and available for exploitation by a MEC node. It may, of course, be necessary to increase extant capacities of any of the three criteria (space, power and network), to meet the increase in demand by the MEC node.

Note that TR-101 "neither requires, nor precludes subtending architectures based on Ethernet transport to remotes" (i.e., remote access nodes, such as are found in some outdoor cabinets). This provision allows distribution of "the complexity of Ethernet Aggregation between the elements of the Access Network and Access Nodes themselves". The significance of this provision is that the task of aggregation hinges only upon the facility of Ethernet transport, rather than the global descriptor of equipment's functionality. For example, both the ONU ***and*** the OLT can meet the requirement of aggregation over an Ethernet uplink.

c) Active distribution sites: good MEC nodes

Of the available real estate, it is clear that ***active distribution sites are good candidates for locating MEC nodes***. To some extent, they possess space, power and network; this is a better start in the attempt to meet criteria than space alone. All contexts of Case B, whether in the so-called far edge or outright in the customer's premises (the FTTB/MDU/MTU, or even customer edge) are such good candidates. This observation brings us to relate MSOs to traditional telcos, i.e., those that have their origins in the public switched telephone network (PSTN) service.

2) MSOs have an advantage from legacy

While telcos are rooted in the PSTN, MSOs are rooted in the coaxial cable distribution network for video delivery (CATV). This was displaced by the HFC distribution network, which reduced the powered points to sites past the HFC architecture's optical node. "N+5" is common in such outside plant (OSP), i.e., 5 powered points past the optical node. Adding to this availability of power all along the coaxial portion of the OSP, is the practice of using a lower impedance power feeder cable that runs parallel to both the fibre and the coaxial distribution. One estimate, dated 2017 [193], is that 80% of the distance covered by HFC OSP in North America, has power supply, and that "in most cases", there is enough power to meet the demands for 5G's small cells. This contrasts with the passive distribution (case A) which telcos have embraced in their migration to PON ODNs.

5.5 Recapitulation

Investigation has shown that the S RP may be counted on as an external (as opposed to subsumed) interface. While this is useful in cross-context analysis, the S RP lies within the customer's domain, and therefore the problem of distribution of energy burden arises if this RP were to demarcate analysis of energy consumption. Moving upstream, the energy analyst may need to carry out a detailed investigation of the functions embedded within equipment near the subscriber's end. The U RP is exposed in current generation access technologies like xDSL and HFC, but is likely to be subsumed in PONs. The advantage of bounding an analysis of energy consumption of access networks at the U RP emerges when it excludes equipment in the customer's premises.

Furthermore, the implementational model may be developed into a physical model through modelling of (a) components in all layer networks involved in transport and (b) that of energy loss in optical and copper media. Supporting architectural constructs at the media layer are defined in [194], and can form part of the bases of such a development.

It is encouraging to note the observation carried in [195, p. 9], that "[m]uch more could/should be written about the reference points between the core network and other networks". As least, as far as the energy analyst is concerned, much remains to be done to support homogeneous reporting. This chapter presents a detailed articulation of the gap. It then proceeds to fill the gap through a two-pronged approach that populates an **`implementational model`** with **`topological components`** that represent current technology. The chapter maps unified representations up to the V RP of the technologies which may be expect to form part of the NGN. The rest of the metro area will be dealt with in Chapter 7.

Chapter 6. Realizing the radical approach: systemic, green virtualization through realizable standards

The radical approach to more sustainable networks employs software defined networking. Most practitioners agree [196] that, minimally, SDN abstracts the data plane, and incorporates a closed feedback loop (thereby supporting real-time adaptation to current and expected conditions). A first reflection on of these requirements will qualify them with other constraints, such as the need to discern key performance indicators (KPIs) of the plant and obtain accurate measurements thereof for feedback. Such reflection will also readily conclude that fine-grained adjustment of the plant's dynamics is superior to coarse adjustment (e.g., turning a plant element on or off). When these considerations are applied to the concern with saving energy, they are necessarily augmented by power control (afforded by the green capabilities) and measurement (which is clearly not trivial with virtual entities[13]). If this much complexity emerges from a first reflection, further study into the detail of implementation of software-defined networking is warranted. This chapter acts on this warrant. A system-level view is explored first, and emergent ramifications are tackled subsequently. It concludes with a reference to a case study on cloud-native video streaming, the details of which are presented in Appendix 7.

6.1 The macroscopic attitude: NFV MANO

NFV Management and Orchestration (NFV MANO) [198] establishes the foundations for further development of NFV within the context of software-defined networking. This claim can readily be made on the premise that it is concerned with a "management and orchestration framework" and with "providing a functional architecture", and notably, with "a description of the reference points between NFV MANO functional blocks and other functional blocks in the E2E [(end-to-end)] NFV reference architecture." It suggests that "[e]ach NFVI-PoP[14] or Administrative Domain can include a Network Controller (e.g., SDN controller) responsible for providing the programmable network interfaces that enable the establishment of connectivity within the domain." Therefore, a proposal to carry out research that includes any aspect within NFV-MANO's scope would be incomplete without integrating the development into its framework and architecture.

Constituent subsystems of the ecosystem continue to undergo intense development by a number of different stakeholders. ***One such subsystem is that which controls use of power***. Two groups of stakeholders in this subsystem are recognizable: those who have invested in the development and

---

[13] The concept represented by the term "virtual entity" is identical to that represented by the term "virtualisation container", as stated in ETSI GR NFV 003 [197]. The two terms are used interchangeably throughout this work.

[14] An NFVI-PoP is real estate within which one or more NFVI-Nodes are housed. This may be a CSP PoP, such as a Central Office, Local Exchange or a Head End, a carrier-neutral datacentre, or some similar physical premises within the landscape of commercial telecommunications.

adoption of the Advanced Configuration and Power Interface (ACPI), and those who have invested in the Green Abstraction Layer. The following sections review the integration of these two technological frameworks for power control, within the framework of NFV MANO.

6.2 Rubber hits ground: the nuts and bolts of the ACPI

In the context of COTS, as well as that of commercial hypervisors, power management is implemented upon the foundations of ACPI. Within the ACPI specification [199], operating systems that manage power through ACPI are referred to as comprising "operating system-directed configuration and power management" (OSPM). Indeed, the power management model is commonly referred to as OSPM/ACPI; this conjoining of acronyms is a helpful reminder that ACPI supports the operational paradigm of ***OS-directed*** power management. This supersedes earlier power management models, where platform firmware had an active role, power management interfaces were proprietary and power management settings were not necessarily consistently organized. Both the hypervisor as well as the general-purpose OS (GP OS) implement algorithms, within structural components referred to as ***power controllers***, that exploit the data structures defined in ACPI, e.g., G-states (global system power states, see Fig. 28), C-states (processor power states) and D-states (device power states). One possible goal of such power controllers might be to maintain full availability, i.e., the device appears to be always on to its end-user but under idle condition, transits to a standby or lower-power state – provided that the latency of the reverse transition is undetectable. Power states address depth of sleep: the zero suffix regards the active, or awakened state, but higher-valued suffixes regard progressively deeper sleep as more sub-systems of the computer are turned off. For example, G3 indicates a mechanical interruption of power supply.

ACPI also describes power-performance (trade-off) states; while power states address depth of sleep ACPI P-states are convenient, general representations of (power-performance trade-off) hardware-specific states of the "on", or "active", operating mode for processors (the C0 state) and devices (the D0 state). Processor-core state C0 – Px states are implemented through voltage- and frequency-scaling. As regards devices, other examples are given in the ACPI specification [199, p. 36], including power-performance adjustments for graphics components, hard disk drives, liquid-crystal display (LCD) panels and DRDRAM (Direct Rambus Dynamic Random Access Memory). System device power-performance states D0 – Px are both implementation and device class dependent (again, see [199, p. 36] for examples of dependency on device class). The P-states, like the sleep states, are the basic states of power management which a control algorithm can adopt in its state-machine management. They are basic in the sense that they are the most granular – and therefore indivisible – conditions of the device which the controller can aim to reach. The term ***primitive*** is used in the standard describing the Green Abstraction Layer to refer to these basic power (***standby***) and power-performance (***power-scaling***) states [200, Sec. 6]. Each such state, augmented with other data pertinent to the GAL's framework, is referred to as a ***primitive sub-state (PsS)***.

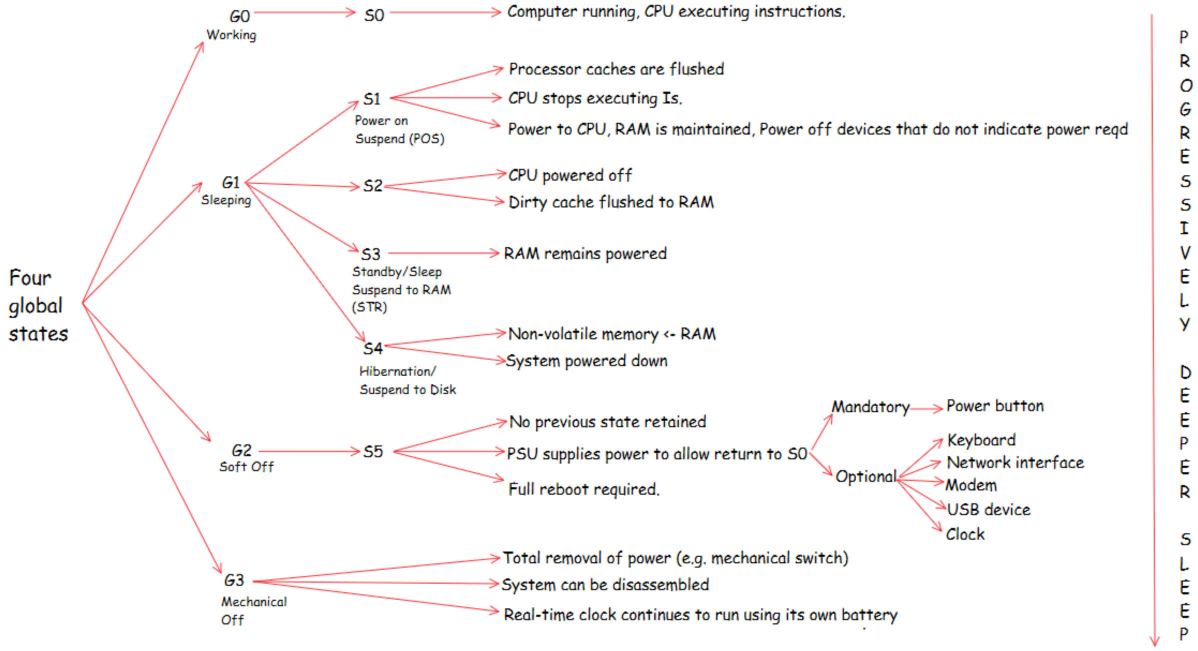


Fig. 28. ACPI global states. G0 is a group of states. Each state is a unique performance-power point.

For the purpose of this section, a power controller with local system scope, will be referred to as a local power controller and denoted by LPwC. This is done to distinguish this entity, within ACPI, that implements a policy for power control, from one within GAL. GAL denotes a similar entity by LCP (local control policy – more precisely, a process that implements a policy for local power control), but since the two entities are ontologically distinct, different terms are used to avoid confusion. In [200], there is no specification of the detailed organization and function of the LCP, but the LCP must interact with GAL-defined data structures; therefore, an LPwC can be recognized as a (potential) component of an LCP.

Within a COTS computer system, two LPwC levels are distinguishable (see Fig. 29 [199]).

1. A level 0 LPwC (LPwC-0) is the lowest-level software entity, one that drives a hardware component (the term "driver" is more recognizable than "controller" at this level). Platform firmware is not within LPwC-0 scope: the ACPI subsystem is an intermediary between platform firmware and system software. ACPI-compliant systems should not use platform firmware for configuration or control. Fig. 29 isolates platform firmware from interaction with ACPI and system software (platform firmware is reserved for boot-time functionality, after which it hooks onto system code).
2. A level 1 LPwC (LPwC-1) is a system software power controller, e.g., a hypervisor or operating system power controller. The interface between level 1 and level 0

controllers may be defined by operating systems, e.g., Microsoft Power Management Framework (PoFx) [201].

3. ACPI defines the data structures of the interface (see the orange zone of Fig. 29)
    a. between LPwC-1 and platform hardware, and
    b. between LPwC-0 and platform hardware (excluding “existing industry-standard register interfaces”, which predate the use of ACPI).

The relationship between the two levels is illustrated in Fig. 29, through the medium of the OSPM /ACPI global system architecture [199].

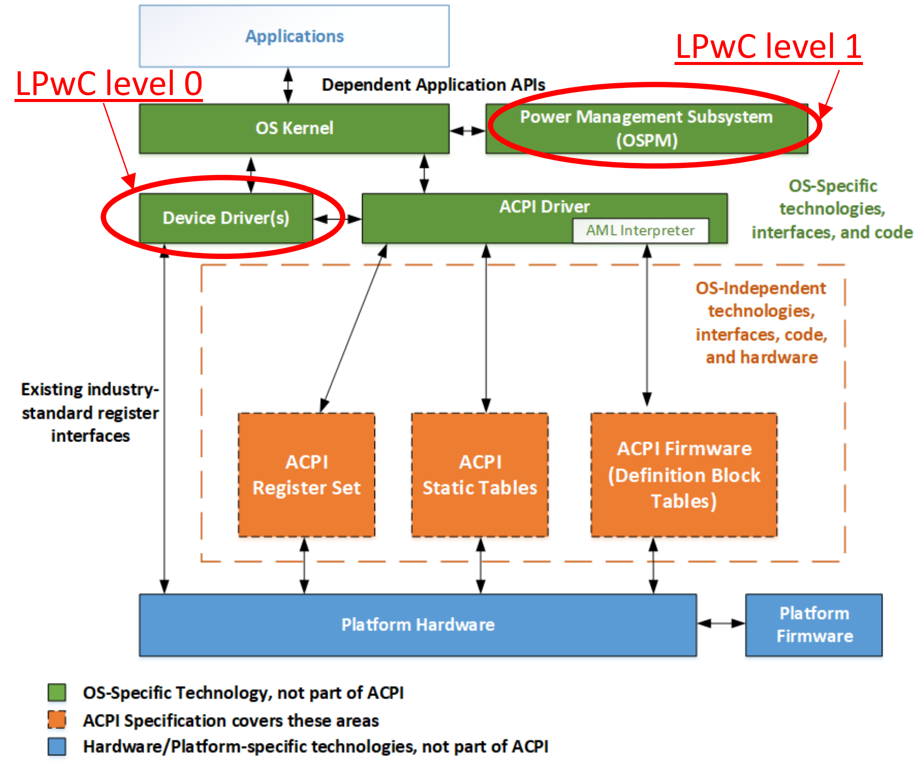


Fig. 29. Power controllers superposed on the OSPM/ACPI global system architecture [199]

Hierarchical LPwCs are compatible with the framework of an ACPI, which proposes, among others, power states that are specific to the elements of a processor hierarchy (see [199, Fig. 8.7]). This intrinsic support provides a valuable interlock to the Green Abstraction Layer’s own use of hierarchy.

6.3 The Green Abstraction Layer has a top-down view

GAL is a reflection of the importance of standardized architectures (see point 2, above) for the concern with bringing the control plane – data interface to bear on sustainability and operational expenditure. A concrete step forward was taken in the year 2014 with the standardization of a green interface, which was embedded within this so-called Green Abstraction Layer (GAL) [200]. A second step extended GAL to include virtualized network functions in scope [202]. Further development is currently underway in Study Group 5 of the European Telecommunications Standards Institute (ETSI).

6.3.1 A philosophy for distributed power control

The philosophy underpinning the Green Abstraction Layer may be summarized in two tenets.

1. Distribute all the power control that you can, and only centralize what you must.
2. Adjust power to just enough to meet the current load's demand for it.

The first tenet leads to a hierarchical architecture comprising a minimalist network control policy (NCP), implemented on a centralized controller and a number of instances of various types of local control policy (LCP). The latter are implemented on each network node (whether hosting VNFs or PNFs) in the GAL domain, where domain refers to the set comprising the controller and all network nodes with which it is authenticated (as an authorized controller). Centralization of governance has the significant disadvantage of increasing (worsening) response times; indeed, this was one of the lessons learnt from the early OpenFlow models of control, when switching devices are reduced to pure data-plane agents (see [203], [204, Ch. 5]).

The second tenet has implications on both the NCP agent and the LCP agents. It may be succinctly expressed as "load-proportional power control". LCP agents are engaged to optimize the configuration at the GAL-device level, in such a manner as to achieve the desired trade-off between power and performance capable of handling current traffic load. The NCP agent is engaged to optimize the configuration at the network level. Through knowledge obtained from an artefact such as a traffic engineering database (TED), the NCP agent selects a path through the network that meets service level objectives, under current traffic conditions, while minimizing power.

6.3.2 An architecture for distributed power control

A simplified view of the major GAL components is shown in Fig. 30. Fig. 30 shows the external (SDN) controller interacting with a GAL-device[15] through the Green Standard Interface (GSI) API. A high-level classification of the API is shown in Fig. 32. The simplified view of GAL architecture may be articulated as follows:

1. GAL-compliance is predicated upon the presentation of the GSI and the implementation of Energy-Aware States (EASs) [200].
    a. An EAS is a complex data type that describes the performance – energy trade-off. Configurable items that enable the operating state characterized by the trade-off are organized into another complex data type, which links to an EAS variable, as a means of establishing a relationship between the two.

---

[15] The GAL standard refers to network nodes as devices. Since this use of "device" poses problems of interpretation when used within the same text where reference to the ACPI's device is made, then GAL-device will be used to refer to "device" where it is to be interpreted as that of the GAL standard.

    b. The EAS is an attribute of an Energy-Aware Entity (EAE). An EAE is an abstraction of some artefact that supports performance – energy trade-off. The artefact is represented in software by a universally unique identifier (UUID) [205] – type variable, denoted `resource_id`.
    c. An EAE possesses at least one EAS (but two EASs would permit state changes according to performance demand).
2. The GAL-device implements a convergence layer (CL), consisting of a GAL-device-specific mix of application-programming interfaces (APIs) and configuration and control registers.
    a. For example, ACPI may be part of the GAL-device's system software.
    b. The GAL accesses the convergence layer through a convergence layer interface (CLI).
    c. The CLI transforms the GSI's abstractions into concrete steps that are in turn transformed by the convergence layer into actions and responses on the underlying platform.

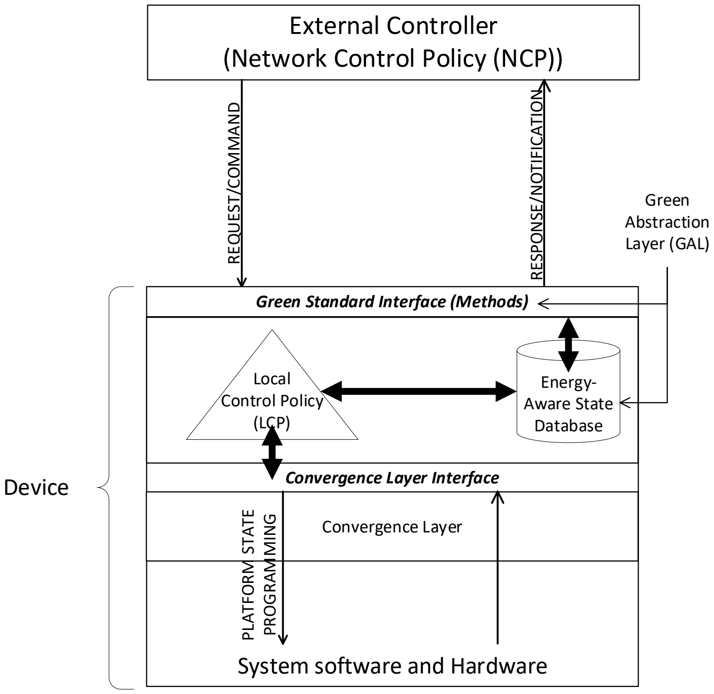


Fig. 30. A high-level view of major architectural components in a GAL-compliant NFV infrastructure (NFVI) network node

Fig. 31 [206] carries a more complex example of how a GAL hierarchy might be constructed, regarding a board (NetFPGA) suited to research and development. Instead of the single EAE shown in Fig. 30, this system may be sub-divided into at least two major branches (the IC and the ports) that have independent green capabilities[16]. The FPGA IC's (field programmable gate array integrated circuit) frequency of operation can be adjusted to match load; its use of power (and thus, energy consumption) can, in this way, be adjusted. The individual ports can be turned into the mode of operation known as "low-power idle" (LPI). In this mode, when a port is idle, its power can be reduced. The overarching abstractions on the group of ports and on the entire GAL-device can capture these facilities and introduce group operations, such as switching all four ports in the group to LPI mode.

Fig. 33 (also from [206]) shows a practical example of a yet more complex GAL-compliant system: a classical network element (the GAL-device), comprising several chassis within a rack. This might consist of a layer 3 IP router, with MPLS, client interfaces, transponders/muxponders and (reconfigurable) optical add/drop multiplexers (IOADMs). Such a GAL-device might include the IP/MPLS functionality within a single chassis, equipped with line cards carrying client interfaces. Another chassis might carry the optical networking equipment (the transponders, muxponders and ROADMs). All chassis would be housed within the GAL-device's rack. The illustration carried in Fig. 33 is particularly useful, as it helps to relieve the essence of each GAL entity. The simplified articulation may now be complemented with a more detailed understanding of the GAL entity.

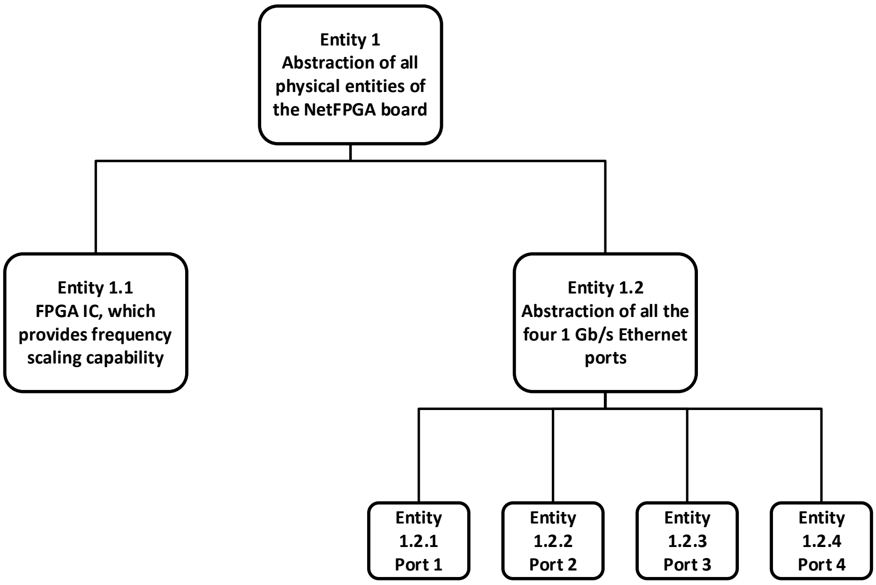


Fig. 31. How a GAL hierarchy might be implemented for a NetFPGA board

1. Each GAL entity is a software object that encapsulates all the (exposed) functionality of a sub-system or component in the GAL-device.
2. The GAL entity consists of:

---

[16] In the GAL context, green capabilities are broadly classified as standby and power scaling. Low Power Idle (LPI) is also used as an alternative term for "standby". Adaptive rate (AR) seems to be more commonly used to describe power scaling in the context of transmission links.

   a. a Green Standard Interface (GSI), through which it interacts with a higher-level entity in the control hierarchy, to implement a control policy (described in more detail further on in this list);
   b. a Convergence Layer Interface (CLI), through which it interacts with specific artefacts of power control (e.g., software/firmware drivers/agents, hardware configuration and control registers). The CLI is software that embeds knowledge of how to use the software/firmware/hardware that embodies the power control. The layer of drivers and hardware-intrinsic software must converge the diversity of the hardware with the homogeneity of the GSI; this is the reason why it is called a convergence layer. In essence, the convergence layer is wrapped by the GSI and the convergence layer then receives a transformation, through the CLI, of the intention of the GSI command into a sequence of hardware actions appropriate to the hardware being driven.
3. Complementary to the GAL entity, ***but not part of the Green Abstraction Layer***, is a Local Control Policy (LCP), which is the code that interprets GSI requests from the higher-level entity, in terms that are used at the CLI, according to some control algorithm, which is in turn guided by the parameters in the GSI requests. Separation of the LCP from the GAL is an intentional omission of scope, to facilitate interoperability with other power management frameworks and architectures.
4. The GSI is the means through which data plane ↔ control plane interaction can take place. Two saliencies which have been used to describe the GSI: that it is a "northbound" interface and that it is "lightweight".
   a. It is ***northbound*** because it is intended as the means for an LCP (at any level) to communicate with a higher (and hence northwards, or towards the control plane) element in the control hierarchy in a manner that frees the higher elements from organizational knowledge of data-plane elements. A brief digression is in order to clarify the architectural awareness of the LCP. At any given level X, the LCP, say, LCP_X is only pre-built with awareness of the logical and physical resources in its own EAE. Indeed, LCP_X may be unaware of any physical resources at all [200, p. 14], as happens when the EAE's architecture exposes no power control. Furthermore, LCP_X does not drive any physical resources at X directly, but only through the CLI. Interaction at levels below X (in the control hierarchy) with hardware resources, takes place through an intermediate GSI implementation. As regards the logical resources, these are software-only elements that serve as a means to aggregate "one or more resources (physical or logical) that provide a summarized and simplified view of the managed entity" [200, p. 12]. The logical resource will also have

    one or more EASs. The GSI, therefore, is truly northbound, as it comprises no knowledge of the data plane except that which is provided by its supporting component: the convergence layer interface.

   b. The GSI is also "***lightweight***". It comprises only six commands, divided into discovery, provisioning, monitoring and configuration-management categories. The simplicity of the GSI is visible in Fig. 32, which is a succinct representation of the functionality in the GSI. Referring back to Fig. 33, it is clear that the GSI is a universal interface in the GAL Architecture. Every inter-entity interaction takes place through the GSI.

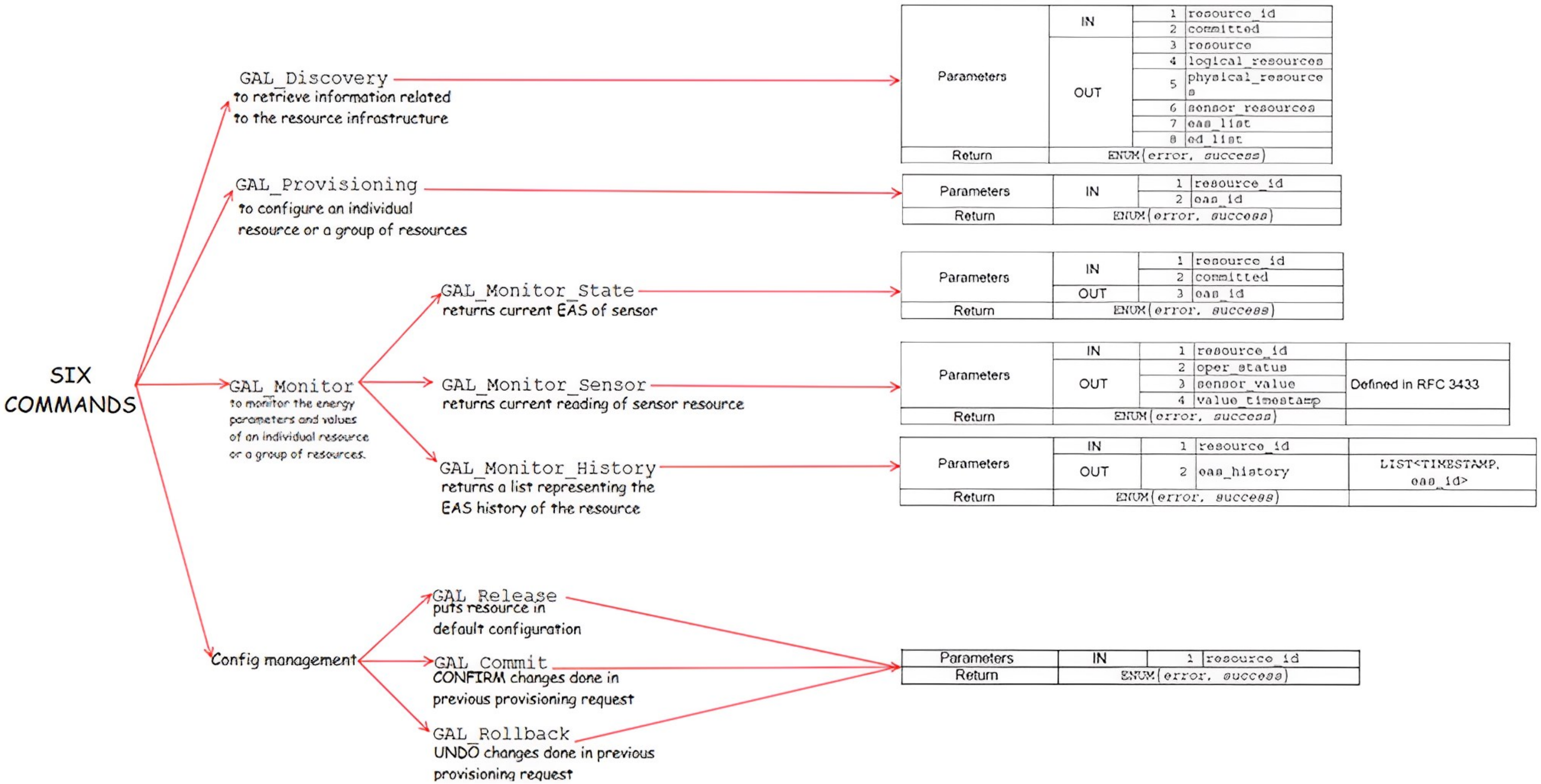
5. The EAS is the data abstraction of a component's hardware's intrinsic capability to perform its function at different levels of energy usage and progressively turn sub-components off when not performing its function (without requiring human intervention to resume function). For an LCP to control the component's energy usage, the hardware must expose this intrinsic capability at its interface.

   Semantically, each instance of state is a set of values, where each value quantifies a specific property in a set of properties, $\{P_r\}$[17], ***that includes power***. A state variable S, therefore, is a set of properties: $S = \{P_r\}$. Furthermore, $\{P_r\}$ must also include performance metrics relevant to the function; from the outset, the philosophy underlying development of the GAL has been load-proportional power control. Henceforth, "instance of state" will be abbreviated to "state", and the generic representation thereof will be referred to as the state variable. At the lowest layer, the EAS pertains purely to physical resources, and the primitive sub-states (PsSs) are an identifiable ***index*** of physical state. Therefore, at that layer, the EAS can be consistently ***indexed*** by a (GAL-compliant) representation of the {standby, power-scaling} pair of primitive sub-states (PsSes). At the higher layer of the logical resource, the EAS is an artefact of the manufacturer who must create a coherent and consistent relationship between:

   a. on one hand, the power and performance of the sub-system which the logical resource represents, and
   b. on the other hand, the indexing of the EASs which capture the control process's perspective of the power and performance.

---

[17] It has already been indicated that the EAS must include values for various power and performance parameters.



Parameters	IN	1	resource_id
		2	committed
	OUT	3	resource
		4	logical_resources
		5	physical_resources
		6	sensor_resources
		7	eas_list
		8	ed_list
Return	ENUM(error, success)		

Parameters	IN	1	resource_id
		2	eas_id
Return	ENUM(error, success)		

Parameters	IN	1	resource_id
		2	committed
	OUT	3	eas_id
Return	ENUM(error, success)		

Parameters	IN	1	resource_id	
	OUT	2	oper_status	Defined in RFC 3433
		3	sensor_value	
		4	value_timestamp	
Return	ENUM(error, success)			

Parameters	IN	1	resource_id	
	OUT	2	eas_history	LIST<TIMESTAMP, eas_id>
Return	ENUM(error, success)			

Parameters	IN	1	resource_id
Return	ENUM(error, success)		

Fig. 32. The Green Standard Interface's API can be classified into six command groupings (Discovery, Provisioning, Release, Monitoring, Commit and Rollback) [200]

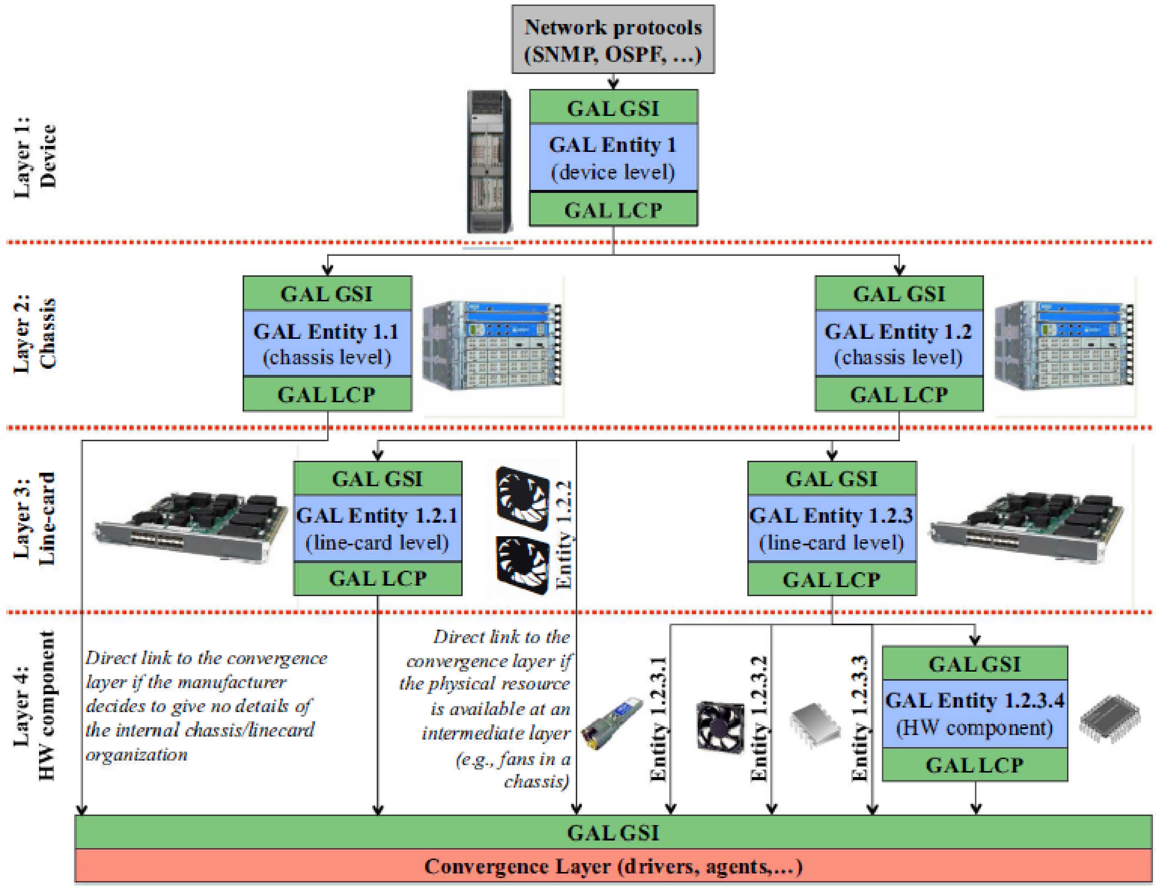


Fig. 33. A GAL-compliant system's hierarchical architecture for power – performance tradeoff [206]

6.3.3 A state space for the EAS

GAL discretizes the EAS's state space through a mapping to standby (sleep) and power-scaling (active) states. GAL obtains a consistent notational description of these primitives through a combination of the active states and the sleep states in the EAS [200, Sec. 6]. The following relationship between state index n and primitive indices j, k is defined:

$$EAS_n = \left\{P_j^{(n)}, S_k^{(n)}\right\} \; 0 \leq j \leq J \text{ and } 0 \leq k \leq K \tag{8}$$

The meaning associated with the (indivisible) primitives is the following [200, Sec. 6]:

1. power-scaling (power-performance) primitive sub-states, $P_j^{(n)}$ (P-PsSes):
    a. sub-state number 0: performance (-relevant metrics) and power use are at their highest and

    b. sub-state number J: performance (-relevant metrics) and power use are at their lowest;
    c. sub-states between 1 and J-1: these represent intermediate performance states while the component is active;
2. standby primitive sub-states, $S_k^{(n)}$ (S-PsSes):
    a. sub-state number 0: the component is active, or “on”
    b. sub-state number K: the system is off
    c. sub-states between 1 and K-1: the component is sleeping and therefore is not performing its function.

The relationship between the index n (a scalar descriptor) in the notation EAS_n and the primitives' indices (j, k), is defined in the GAL standard [200, Sec. 6] as follows. Values of n are presented in ascending order.

- $n < 0$: ***Standby***
    - **Description**: the EAE is in standby, and therefore is not performing.
    - **Range of values of j and k**
        - $j = 0$ is used to denote the standby EASs. Since the power-scaling primitive is meaningless in this case, the choice of value is, more or less, arbitrary.
        - $0 < k \leq K$
    - **Values of n, in ascending order**
        - $n = -K$ represents the EAS with the least power gain, i.e., that with the shallowest sleep (and shortest wake-up time)
        - $n = -1$ represents the EAS with the greatest power gain, i.e., that with the deepest sleep (and longest wake-up time)
- $n = 0$: ***Active, with maximum performance and power use***
    - **Description**: in this state, the EAE is performing best and consuming the most power (and therefore there is no power gain).
    - **Range of values of j and k**
        - $j = 0$
        - $k = 0$
- $n > 0$: ***Active, with sub-maximal performance and power***
    - Description: The EAE is in a throttled mode of operation. Power-scaling primitive sub-states divide the active mode into several rationalized states, each of which saves power with respect to the maximum power and performance state.

- **Range of values of j and k**
    - $0 < j \leq J$
    - $k = 0$
- **Values of n, in ascending order**
    - $n = 1$ represents the EAS with the least power gain and corresponds to $j = 1$
    - $n = J$ represents the EAS with the greatest power gain and corresponds to $j = J$

A single EAE therefore has $\{(K) + 1 + (J)\}$[18] EASs. The maximum number of EAS that an EAE can support is the product of the number of the two PsSs, i.e., $(K + 1) \times (J + 1)$ [200, p. 19], but only $\{K + J + 1\}$ are physically realizable. However, for the sake of extensibility, the GAL standard postulates custom EASs that are inaccessible to external control processes, whether local to the component (LCP) or external and part of its network's control (NCP). These custom EASs are denoted by the combined constraints $j > 0,\ k > 0$. The range of values of n above $n = J$ has been reserved for these EASs (i.e., those that are outside the GAL's control scope).

6.3.4 Modus operandi: fine-grained, distributed power control

Fine-grained, distributed control is effected through a standardized operating procedure (workflow). Consider the case of a centralized network controller which uses an energy-aware network control policy (NCP). The NCP determines that a logical link (at the IP layer, say) can be brought down. A procedure must be devised for sending the relevant port, on the relevant line-card, in the relevant chassis, in the relevant GAL-device, to sleep. A data-flow diagram (DFD, Fig. 34) is used to show the interaction between the hardware and software components. The numbers on the data flow arrows show the order in the sequence when the data flow takes place. The overall architecture implicit in Fig. 34 is consistent with that shown in the workflow illustrated in [200, Fig. B.1]: the GSI, "GAL Library" and LCP interact as indicated in that workflow. Note that a DFD does not need to follow (what ITU-T Y.110 calls) the implementational model, i.e., it does not need to show functions as they are placed in their containers. Indeed, the DFD in Fig. 34 lumps all instances of the GSI into one entity. It is the exchange of communication that is being modelled here.

GAL_Provisioning configuration commands are shown in steps 1, 3, 5, and 7. This reflects a progressive reduction of scope from the abstraction of link to the specifics of port and eas_id. The eas_id, in particular, changes according to the EAS at the level it pertains to. After step 7, there is no

---

[18] The EAS actually has $\{(1 + K) + (1 + J)\}, = \{K + J + 2\}$ PsSs.

further depth to drill down to; the convergence layer interface can be invoked through a wrapper method (step 8) that indicates the standby primitive sub-state desired. Once the CLI obtains the desired change, a sequence of state transitions unfolds: steps 12, 15, 18 and 21 regard EASs that are progressively higher summarizations of logical and physical resources lower in the hierarchy. Each such transition is followed (steps 13, 16, 19 and 22) by a return status code indicating success. The event by which the CLI notifies the port's transceiver's LCP is shown as step 11.

Note that apart from the port-EAS, the line-card-EAS (and the chassis, and the GAL-device-EAS) is modified too. This follows because the line-card-EAE's EAS is a summarization of the lower-layer EAEs' states. Therefore, a change in state at a lower level will, in general, affect the EAS of the "parent" EAEs.

6.4 GALv2: Standardized support for virtual entities

ETSI Standard ES 203 682 [202] extends the Green Abstraction Layer standard [200] (and calls itself GALv2) to support "power management in Network Function Virtualisation (NFV) environments". This update is concerned with "establishing a mapping between the EASs of logical entities (e.g., VNFs) and the energy consumption of the hardware hosting the virtual machines". Obliquely, this fragment of a sentence refers to the VNF as a ***logical resource***. This marginal reference is critical in understanding how power-aware performance control is obtained in NFVI. First, however, an overview of the accomplishments of this update, is warranted.

6.4.1 An overview of ETSI Standard ES 203 682

The standard styles itself as the adaptation of GAL specification ES 203 237 to the NFV environment. It defines itself within the context of the ETSI NFV suite of reports and standards (notably [207], [208], [209], [210], [211], [212], [213] and the architectural framework [81]) thereby supporting its own adoption through the use of a globally-recognized foundation. The ETSI NFV suite, employing the terminology established in [197], creates an entirely new ontological plane, the inhabitants of which it weaves together using a epistemology the robustness of which is witnessed by its global adoption. Some of the inhabitants of this new plane readily correspond to what the GAL standard describes as ***logical resources***. However, network function virtualization is not a mere augmentation of the landscape of network functions (NFs) with virtualized network functions. It merits the treatment of a paradigmatic shift, and this includes a detailed investigation of the particular logical resources that it entails. With this in mind, it can be seen that the primary contribution of the GALv2 standard, is application of the key artefacts of ES 203 237 to the new ontological plane. In so doing, GALv2 extends the attributes of several of the NFV architecture's elements, rendering such elements amenable to inclusion within the scope of power-aware performance control. The contributions of the GALv2 standard may be summarized as follows.

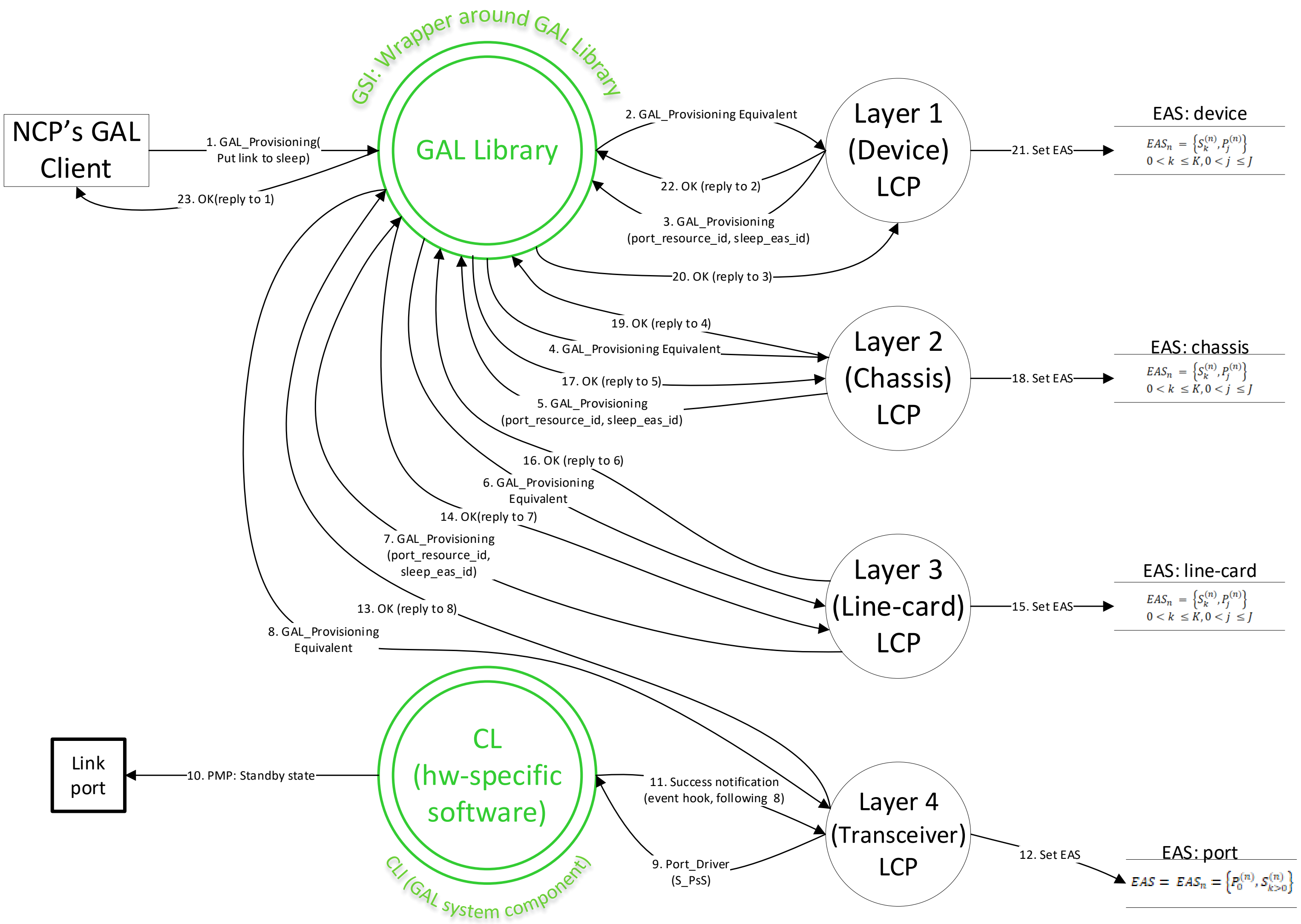


Fig. 34. Data-flow diagram showing interaction between components of the GAL architecture

8)) EAS definition

Energy-aware states are defined for virtualized network function components (VNFCs), VNFs and network services (NSs).

1. VNFC's EAS: The definition hinges upon the virtualization container (VC) on which the VNFC is deployed, but the EAS's indexing rule is identical to that stated earlier. The standard limits itself to identifying standby states and power-scaling states, but does not address how depths of sleep or power level (respectively) are achieved. This problem is addressed later in this work. Physicality (the resources that use power) is obtained through instantiation of the VC on a specific compute domain (CD) and hypervisor domain (HD). See Fig. 35. Identification of the couple (CD,HD) is foreseen through a function that takes the specific VNFC instance as argument [202, Para. 5.2], ***There is no clear reference to the network domain*** (ND) in the standard's exposition of physicality; the standard's text should be updated to clarify whether it is intended that the power use of the ND is subsumed within the other domains, or whether it has been overlooked.
2. VNF's EAS: Since each virtualized network function may comprise several types of VNFC, and each VNFC-type may be instantiated several times, the standard limits itself to defining a notation that captures the n^{th} EAS. ***Thus, the*** $\mathbf{n^{th}}$ ***EAS is a set of (column) vectors, with each vector representing a specific VNFC-type, and the number of elements in the vector representing the number of instances of the VNFC-type***. Consider any one (column) vector. Each element in this vector represents a VNFC EAS, but the elements do not need to be in the same EAS, although the respective elements are members of the same set (consisting of all that particular VNFC's EASs). While this is relatively arbitrary, the scope of further standardization is so vast that this extent of regulation is already useful in establishing the terms of the problem. Furthermore, it should be kept in mind that the physicality of the VNF is nested within the physicality of the VNFC's virtualization container.
3. NS's EAS: This EAS can readily be understood as related to the VNF's EAS in the same manner as the latter is related to the VNFC's EAS. The same pattern holds: a set of (column) vectors, with each vector representing a specific VNF, and the number of elements in the vector represents the number of instances of the specific VNF. As regards physicality, there are now two levels of nesting to traverse before the physical compute, storage and network elements that

   Extension of scope of the EAE to NSs is novel to GALv2. With GAL, the highest entity in the hierarchy is the "device", and a multi-chassis, rack-mounting node is used as an example. With NSs as EAEs, GALv2 limitedly formalizes the data structure which is the object of optimization, extending thereby into the realm of the control plane.

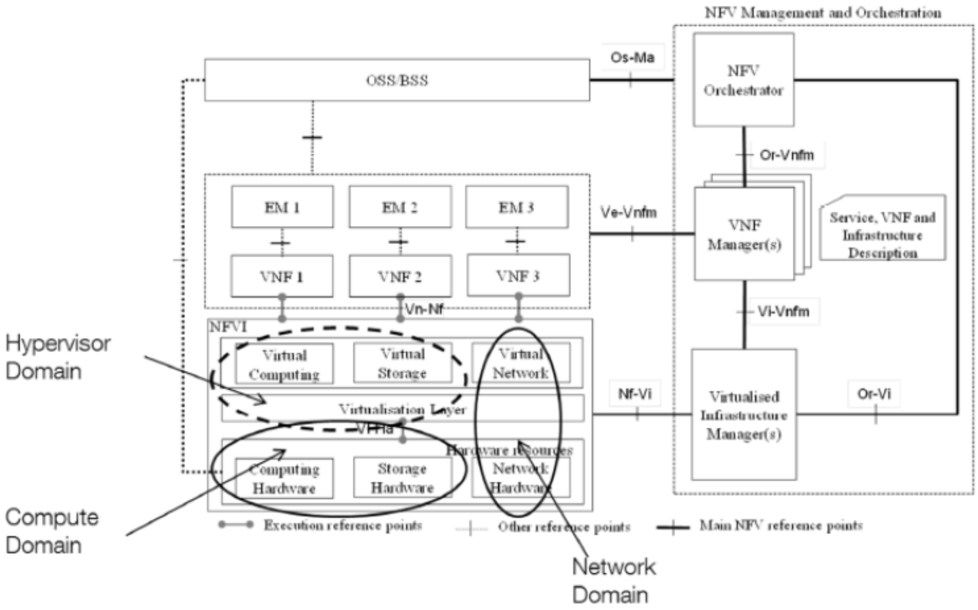


Fig. 35. Superposition of compute, hypervisor and network domain on the NFV-MANO architecture [202, Fig. 4]

2) Mapping GSI to NFV-MANO

The mapping can be mathematically expressed as:

$$f\{\mathrm{GAL\_Provisioning}, \mathrm{GAL\_Monitoring}, \mathrm{GAL\_Release}\} \\ \rightarrow \{\mathbf{I}(\mathrm{Os}-\mathrm{Ma}-\mathrm{Nfvo}), \mathbf{I}(\mathrm{Or}-\mathrm{Vnfm}), \mathbf{I}(\mathrm{Or}-\mathrm{Vi}), \mathbf{I}(\mathrm{Vi}-\mathrm{Vnfm}), \mathbf{I}(\mathrm{Ve}-\mathrm{Vnfm})\}$$

where $\mathbf{I}(\mathrm{Os}-\mathrm{Ma}-\mathrm{Nfvo}), \mathbf{I}(\mathrm{Or}-\mathrm{Vnfm}), \mathbf{I}(\mathrm{Or}-\mathrm{Vi}), \mathbf{I}(\mathrm{Vi}-\mathrm{Vnfm}), \mathbf{I}(\mathrm{Ve}-\mathrm{Vnfm})$ refers to the interfaces at the indicated NFV-MANO reference points. The details of the mapping are superfluous to a summary, but useful observations can be made that convey the underlying rationale.

1. In so far as management and orchestration are concerned, the GALv2 standard complies strictly with the NFV suite of standards and reports. Consider the following examples of compliance.
    a. The origin of a GALv2 – compliant network service (NS) lies within the operations support system / business support system (OSS/BSS). The OSS/BSS communicate with the NFV Orchestrator (NFVO) across the Os-Ma-Nfvo RP, to obtain instantiation through the orchestrator.
    b. Control flows down a hierarchy. The NFVO interprets the network service descriptor (NSD) and invokes the virtualised network function manager (VNFM) to instantiate VNFs. The VNFM invokes the virtualised infrastructure manager (VIM) in the process of executing the NFVO's request. Subsequent modifications to the state of the VNFs, e.g., starting them, scaling them and modifying their metadata, follows the same hierarchy of control.
2. GALv2 embeds within the NFV suite through use of the suite's structural provision for expansion parameters in the messages defined by the various members of the NFV suite. For

example, when a human operator drives the OSS/BSS to instantiate a network service, the `InstantiateNsRequest()` takes an `additionalParamForNs` parameter, and the GALv2-compliant OSS supports identification of a specific EAS. Thereby, upon startup, this high-level state is obtained. Evidently, this EAS must be transformed in the process of instantiation that follows the downward flow of control from OSS/BSS to VIM.

3. GALv2 supports closed feedback through data-plane driven `Notify` messages. For example, note the structural support to feed an event (possibly aggregated with other events) through from the data plane (VIM) right up to the highest level of control (OSS/BSS):
    a. VIM→VNFM: `PerformanceInformationAvailableNotification`, `ThresholdCrossedNotification`
    b. VNFM→NFVO: `PerformanceInformationAvailableNotification`, `ThresholdCrossedNotification`
    c. NFVO→OSS/BSS: `PerformanceInformationAvailableNotification`, `ThresholdCrossedNotification`

6.4.2 Dynamics of control in GALv2

Currently, the GALv2 standard carries informative annexes which show outlines of the Provisioning, Release and Monitoring operations ([202] Annex A - C). A configuration flow is also shown ([202] Annex D). However, there is no indication of which component within NFV-MANO might play the role of local control policy (LCP) process. The two standards' (GAL and GALv2) approaches can be merged through insertion of LCP processes at the NFVO and the VNFM. Fig. 36 shows how the VNFM LCP might react to reduce power use on load reduction.

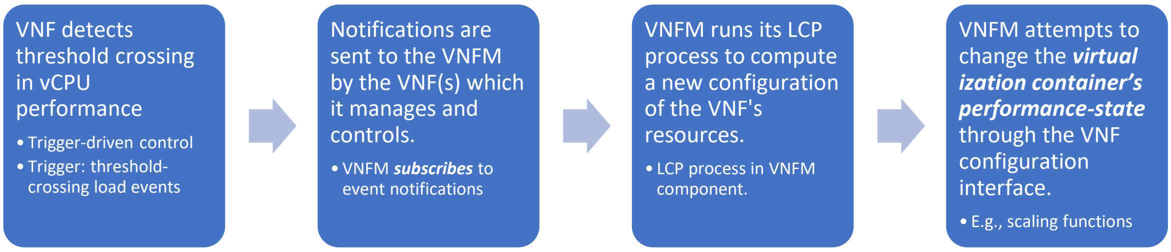


Fig. 36. LCP process in VNFM effects power control through the support of notifications

6.5 An architecture for NFVI-Node power control predicated upon ACPI and GALv2

6.5.1 System components and schematic

An attempt to design an architecture that combines ACPI and GAL would do well to take their respective scopes into account. ACPI's role lies squarely within what GAL refers to as the convergence layer; this is very pointedly expressed in Fig. 37 [199], which places the ACPI subsystem as system software's standardized intermediary to platform firmware. For widespread adoption, GAL must not supplant ACPI within COTS systems, as manufacturers of such systems have high vested interest in using ACPI as the means through which OSPM is obtained (for motherboard, as well as add-on boards).

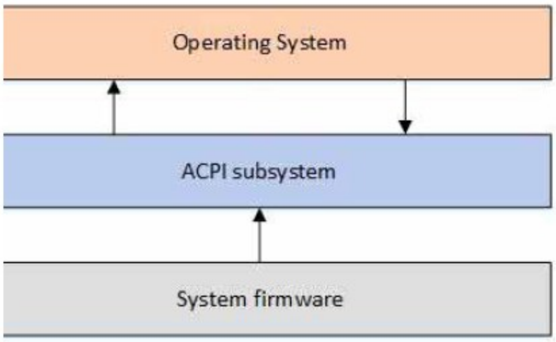


Fig. 37. The ACPI subsystem's location within GAL's CL emerges from its relationship to major subsystems in COTS hardware [199]

In similar mode of thought, GAL fits well in the higher levels of a GAL-device control hierarchy. Compliance is predicated upon an interface (the GSI) and a data structure (the EAS), both of which can ***extend*** existing technologies. This is a good principle for supporting adoption; the converse, i.e., forklifting technology layers (or, worse still, stacks of layers), in favour of new ones, may well occlude any technical benefits that a novelty carries. Therefore, a first sketch of GALv2+ACPI power management of a COTS system would show the following layers:

1. The CL would consist of the ACPI in the lower sublayer and the device drivers (LPwC-0) would comprise the upper sublayer.
2. The API to the set of the device drivers and the ACPI would jointly comprise the CLI.
3. The GAL-device's external interface (facing the higher-layer control agent) would be a GSI. GAL compliance requires that the network controller – network device interface (where the network device is the NFVI-Node) be GSI. Therefore, an instance of an LCP must exist here (since a GSI is associated with an instance of an LCP, see Fig. 30 and Fig. 33).

Proceeding further: on an NFVI-Node, it can be anticipated that the highest-level LPwC will have ***a negotiating role***. This follows because this LPwC is the intermediary between two or more parties – one or more tenants (customers who run virtualization containers (VCs) on the NFVI-Node) and the infrastructure provider (InP, equated with “network provider”[19] in [214, p. 67]), who operates the NFVI-Node. VCs (tenants) compete for the same resources to guarantee individual performance requirements. Behaviour of each party in terms of power and performance is not independent of the other parties’ behaviour and compromises may have to be sought. The NFVI-Node’s power capacity is limited, and NFVI-Node EASs that meet all parties’ requirements for performance may not exist. Therefore, the highest-level LPwC must include ***a broker*** for the VCs hosted on the NFVI-Node. It is a broker in the sense that it can be anticipated that this LPwC will negotiate VCs’ power requirements with infrastructure providers’ power requirements and power capacity. Some data store must exist, comprising VCs’ current EASs. It is useful to conceive of this as an EAS-broker function within the role of this highest-level LPwC. Rather surreptitiously, the role of the LPwC has started to merge into the role of the LCP. If this highest-level LPwC is interacting with the EAS database, then it must at least be GAL-aware and, at least in this limited sense, conform to the GAL architecture.

One candidate for this expansion of role from LPwC into LCP is a system software manager that converges all interactions with power-aware hardware. This LPwC must conform with extant, standardized hierarchy. Clearly, ACPI supports hierarchical control: earlier, it has been seen that hierarchical disposition of control is obtained within ACPI. A clear articulation of the separation of control scope is found in Microsoft’s implementation [215, p. 595]. Microsoft’s implementation, referred to as “power manager”, (i.e., the one described in [215]) serves as an example of the potential for extant COTS systems to expand beyond ACPI-compliance and GAL+ACPI – compliance. In such an effort, “power manager” would correspond to LPwC-1, described earlier, and would be expanded to post status and receive commands through the EAS database.

However, a serious objection to this approach might relate to the impact on system stability of significant extensions to code that is in use. In this case, a third level of local power controller (LPwC-2) might be used. ***This would be the GALv2-compliant entity***, and since it lies above the system power controller (LPwC-1), it would be at the highest level in the hierarchy of power control on the GALv2-device, i.e., it is the layer 1 LCP (see Fig. 33 for clarification of layers in GAL). The LCP would be expected to include at least the four modules described below (the following references to the term ***device*** use it in the sense defined in the ACPI specification [199, p. 16], i.e., ***a constituent element*** of the system – unlike the GAL standard, which uses device to refer to the system). A description of the EAS database record is appended to the list describing the four modules, as the fifth element in the list.

1. The **GSI Listener & Dispatcher** listens for incoming requests/commands and interprets them, invoking other modules, as necessary, for further processing.

---

[19] From [214, p. 67]: “Network Provider. Sometimes also called the Infrastructure Provider (InP) or wholesaler. ***Owns and operates the physical access network***.” (bold and italicized style is my addition)

    a. It interacts with higher layers in the control hierarchy to support:
        i. the GSI's Provisioning commands, by interacting with the EAS broker, such as when a VC requests a change of its EAS;
        ii. the GSI's Monitor commands, by interacting with the EAS Broker (`GAL_Monitor_State`), the Green Tuning CLI (`GAL_Monitor_Sensor`) and the Green Policy Maker (`GAL_Monitor_History`), and
        iii. the entity configuration management commands (`GAL_Release`), by interacting with the Green Policy Maker.

   GSI output parameters and return codes (error, success) are also handled by this module.

2. The **Green Tuning CLI** interacts with the convergence layer to execute ***platform-specific device-level power-performance*** related activities. The Green Tuner executes device-level adjustments, by calling CL code that keeps a device on but changes its capacity (in whatever unit is pertinent to its capacity, e.g., IOPS (input/output operations per second), MIPS (millions of instructions per second), bps (bits per second), etc.). This module's scope straddles the LCP – CLI boundary.
3. The **EAS Broker** evaluates the feasibility of the requested change of state, and executes the change of state where feasible. Feasible cases are transformed into steps that correspond to power primitives and execution is handed over to the Green Tuning CLI, for fulfilment on a green-capable component, e.g., to change a processor core's frequency of operation. Examples of infeasibility include:
    a. the case where it is not be possible to execute global power state transitions with immediate effect, and the invoking control element would need to be informed (through the Green Listener and Dispatcher);
    b. in multi-tenancy, this module must balance the demands of multiple tenants with the InP's capacity to supply power and remove heat. If, for example, a VNF requires a higher processor performance state (P-state) but system temperature is above a safe threshold, then the EAS Broker might decline the request.
    c. Another example might concern a physical network port. If a VNF is turned off, the EAS broker determines whether it is safe to change the physical port's EAS state to some deep standby state (e.g., D3hot), or whether it is mapped to some other VNF on the NFVI-Node (which may prevent any power state transitions).

The EAS Broker is the only module which can interact with the EAS Database.

4. The **Green Policy Maker** has a supervisory role within the LCP. It is the LCP's control and management core.
    a. The control core executes algorithms with a node-local perspective on optimization of power use. It interacts with the Green Tuning CLI, to drive actions and receive feedback from the platform about power state, temperature and other related data. In this manner, the Green Policy Maker serves as a single point of reference, transforming the Monitoring commands into intra – LCP calls to the Green Tuning CLI.
    The extent of control which the Green Policy Maker implements is perhaps the most arbitrary and undefined of all the modular roles' specifications. A minimalist implementation would concern itself only with internal housekeeping, i.e.:
        i. instantiation of GAL's logical resources (logical resources are purely GAL artefacts that abstract aggregations of other resources [200, Sec. 5])
        ii. the control methods necessary to encapsulate the steps inherent to `GAL_Release`.
        iii. ensuring that all other modules are stable and
        iv. that the EAS Database is integral.

    A maximalist interpretation of the role would aggressively pursue the power – performance balance objective, and would largely shift control of power into its court. The boundary between platform-native power control and GAL-native power control is customizable. For example, Windows leaves delegation of idle detection to the device driver. If the driver wants the power manager to undertake idle detection, it calls Windows' power manager function `PoRegisterDeviceForIdleDetection`.

    b. It interprets Discovery within the NFVI-Node's context. It maps devices enumerated by the platform into GAL-compliant resources and assigns a GAL-compliant UUID [205] – type variable (`resource_id`), as follows.
        i. It interacts with the platform's OSPM system code, through the use of methods that permit the Green Policy Maker to obtain information about enumerated devices' individual power-management capabilities, and

    ii. It interacts with the EAS broker such that, for every individual VC, a physical resource tree is created that corresponds to the VC's assignment, and links the VC's EAS database record to the physical resources which are at least partially assigned to it.

   c. It configures the OSPM's system power policy, through interaction with the platform's OSPM system code. Configuration of system power policy may include such settings as configuration of minimum and maximum processor package frequency of operation.

   Since the Green Policy Maker's interacts with the OSPM system code, it partially implements the NFVI-Node's CLI. Interactions are limited to management functions; real-time control is obtained through the Green Tuning CLI.

5. The **EAS database** complements the LCP. The structure of the record is outlined in [200, p. 19] and the recommendation presented here develops it. Basing on that outline, the EAS record may be denoted by $S_{EAS}^{(n)}$, where:

$$S_{EAS}^{(n)} = \left\{P_w^{(n)}, P_r^{(n)}, P_d^{(n)}\right\}$$

   and:

   - $P_w^{(n)}$ is the set of power-specific attributes of state n,
     - e.g., $P_w^{(n)} = \{(Maximum\ Power\ Use)^{(n)}, (Power\ Gain)^{(n)}\}$
   - $P_r^{(n)}$ is the set of performance-specific attributes of state n,
     - e.g., $P_r^{(n)} = \{(Packet\ Throughput)^{(n)}, (Packet\ Latency)^{(n)},\ (Packet\ Loss)^{(n)}\}$
   - $P_d^{(n)}$ is the set of delay attributes of state n
     - e.g., $P_d^{(n)} = \left\{(Delay\ to\ S_{EAS}^{(0)})^n, (Delay\ to\ S_{EAS}^{(1)})^n\right\}$

GALv2's implementation of GSI within the NFV-MANO architectural framework [198, Ch. 5] supports the purported interaction with the GSI Listener and Dispatcher at the Or-Vi and Vi-Vnfm RPs. Therefore, the proposed layer 1 LCP is a component of NFV-MANO's Virtualised Infrastructure Manager. A schematic that illustrates the relationship between the system components, as defined in this sub-sub-section, is presented in Fig. 38.

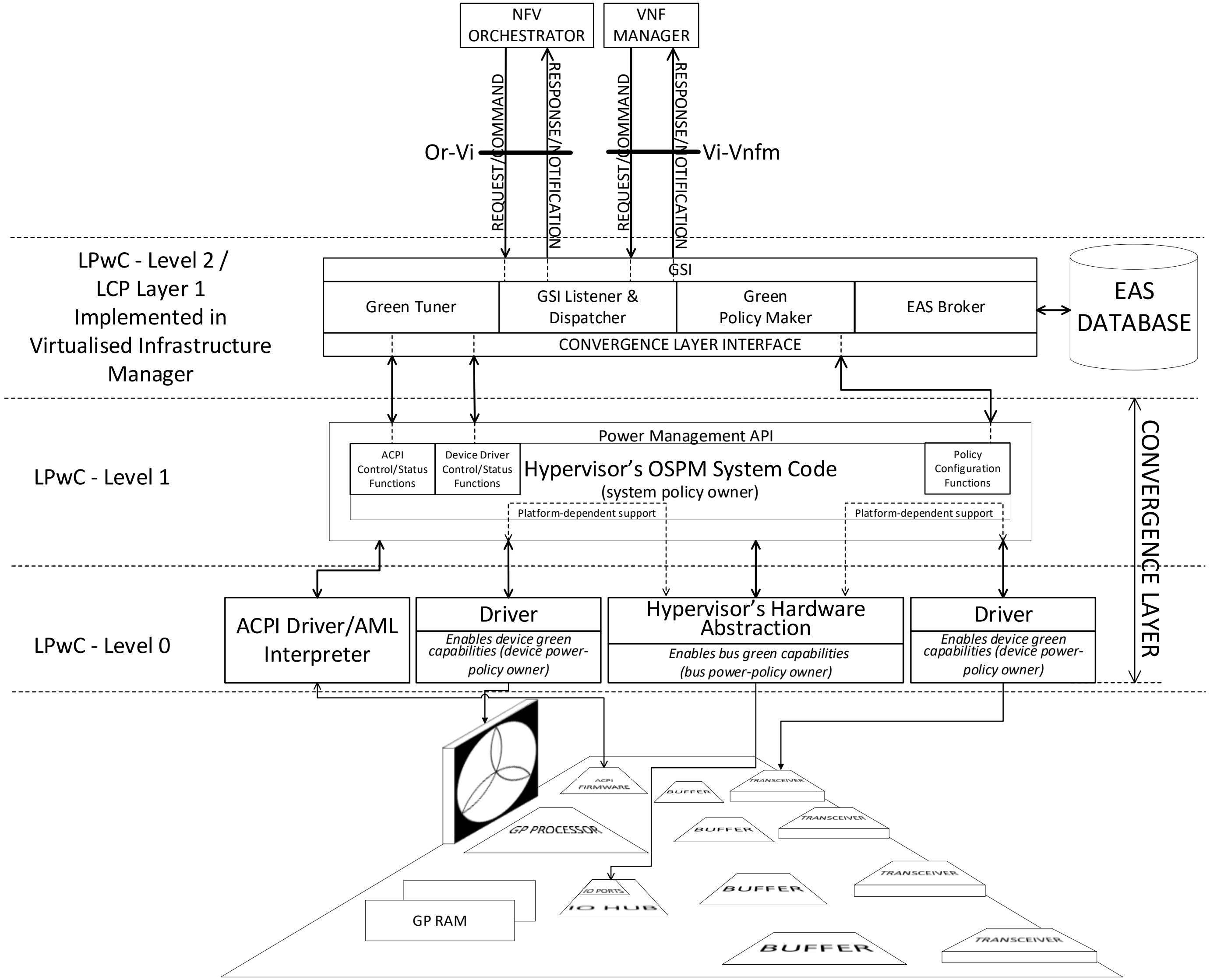


Fig. 38. Architecture of a GAL - , OSPM – and ACPI – compliant COTS computer system platform

6.5.2 Example: definition of the EASs of a logical resource composed of two physical resources

Perhaps the single, most pressing problem that a GAL architect faces is comprehensive definition of the EASs of logical resources. EASs of physical resources are characterized by two primary dimensions: a dimension that describes depth of sleep (or standby) and another that describes level of activity (or power-scaling). If a second physical resource of the same type is added, and a logical resource used to describe the aggregate of the two, then the two-dimensional approach no longer suffices to describe the logical resource's EAS. This might be resolved by increasing the number of states in both dimensions.

Thus: if each physical resource has K sleep states,

then the logical resource would have a maximum of K^2 and a minimum of $\frac{K(K+1)}{2}$ sleep states,

the latter being the case for a logical resource that is insensitive to the ordering of its constituent physical resources (perhaps, but not necessarily, this would be the case when the physical resources are identical).

For the logical resource's active states, the consideration must span all possibilities where at least one core is active.

Thus: if each physical resource has K sleep states and $J+1$ active states,

then the logical resource would have a maximum of $(J+1)(J+K+1)$ and

a minimum of $\frac{(J+1)\left((K+J+1)+(K+1)\right)}{2}$ active states

For example, if the physical resource is a processor and the two processors are in deep sleep, then the depth of the sleep of the logical resource is greatest. If one processor rises to a shallower level of sleep, then the depth of the sleep of the logical resource rises to a shallower level.

This simple approach can be extended to the case where the second physical resource is of a different type. However, as pointed out while discussing discretization of state space, the need to rationalize quickly becomes pressing, as the number of possible combinations forces the question of how many of these possibilities are realizable and useful. Finally, in order to preserve the GAL standard's indexing notation, some criterion must be adopted that sorts these states into an ascending order. Perhaps the index might sort the power consumed in each active state in a descending order, with highest consumption indexed by $n = 0$ and lowest indexed by the highest numerical value.

6.5.3 A generalized approach to defining the energy-aware states of the NFVI-Node

The NFVI-Node is defined in [197, p. 10] as the "physical device(s) deployed and managed as a single entity, providing the NFVI Functions required to support the execution environment for VNFs".

By now, it has become apparent that a GALv2-compliant NFVI-Node include at least the part of the Virtualised Infrastructure Manager that is responsible for the GSI, the LCP and the CLI. This complements the COTS computer system hardware and its firmware, and its system software (notably, the hypervisor. It now remains to address the problem of definition of the EASs.

Control of a component's energy use is predicated upon the availability of EASs. The same holds true for a system of components (of various types and various performance levels per type). The control objective(s) of an LCP are limited by the granularity of its state space. Within this space, the control process drives transitions that seek to achieve the control objective; each such transition is effected by the execution of a control (-plane) action in order to adapt to external conditions. State transitions between states are typically limited; a full mesh between all states is not to be expected (see, for example, the various state diagrams in the ACPI specification [199]). Moreover, transitions are limited by disclosure of architecture: if the manufacturer discloses the latency of the transition, power budget (or power margin with respect to the source state) in the destination state and performance metric(s) (or performance margin(s) with respect to the source state) in the destination state, then the architecture discloses sufficient information to support inclusion of the transition within the overall control scope. This depth of disclosure can be compared to the depth obtained by knowledge of the transfer function of a continuous analogue process: the transfer function is sufficient to effect closed-loop control of the process without knowledge of the process's implementation, thereby protecting the manufacturer's design from unnecessary disclosure. Interest is not limited to the active states (e.g., for a COTS computer system, these are the states grouped under G0); the various sleep states (those grouped under G1 – G3 for COTS computers) are also ***basic states of power management*** which a control algorithm can adopt in its state-machine management.

The state space can be uncountably infinite if the power (use thereof) property (or indeed, any property in S) is treated as a continuous variable. Alternatively, the power property can be rationalized into a countable set through well-defined ***operating modes*** of the component. This approach is exemplified in the ACPI's global states of a computer system (Fig. 28). For example, ACPI G0 encompasses a whole group of states which trade off power and performance. Each state in G0 is represented by a state variable of the form

$$\left(\ \left\{C_l 0\left(\left\{P_1^{C_l}, \ldots, P_{R_{C_l}}^{C_l}\right\}\right)\right\}, \ \left\{D_m 0\left(\left\{P_1^{D_m}, \ldots, P_{R_{D_m}}^{D_m}\right\}\right)\right\} \ \right) \tag{2}$$

where:

1. the braces represent sets of states (of processor cores and devices), as there may be several processor cores and several devices represented in each such state of G0;
2. $C_l 0$ represents the l^{th} power-aware processor-core's (i.e., C_l) operating state;
3. $D_m 0$ represents the m^{th} power-aware system device's (i.e., D_m) on-state;

4. each P-symbol represents a property pertinent to the EAE (core or device), e.g., power use, latency to transfer from a neighbouring state, a performance metric like IOPS, MIPS and bits per second, etc., and
5. R_{C_l}, R_{D_m} represent the number of properties of the EAEs C_l and D_m respectively.

The form shown in (2) starts to map ACPI's ontology to GAL's. The COTS computer system is clearly a logical resource in GAL's ontology and as such, there is no straightforward mapping between the PsSs and EASs. Contrast this with the trivial mapping obtained with the processor (a physical resource), between the GAL's {Power Scaling Primitive sub-State (P-PsS), Standby Primitive sub-State (S-PsS)} and ACPI's {Processor Power State, Processor Performance State}. For the logical resource, each mapping point between PsSs and EASs must carry a full complement of properties (i.e., the set $S = \{P_r\}$) necessary to support the logical resource's local (and higher) control policy processes. There is some support in the GAL standard for this [200, p. 19], where it is mandated that power consumption parameters, network performance indexes and (state-)transition parameters accompany each EAS's {P-PsS, S-PsS} pair. This mandate can be used to improve the notation of the energy-aware state as:

$$S_{EAS}^{(n)} = \left\{P_w^{(n)}, P_r^{(n)}, T_d^{(n)}\right\} \tag{3}$$

where:

- n is the state's index
- $P_w^{(n)}$ is the set of power-specific attributes of state n
    - E.g., $P_w^{(n)} = \{(Maximum\ Power\ Use)^{(n)}, (Power\ Gain)^{(n)}\}$
- $P_r^{(n)}$ is the set of performance-specific attributes of state n
    - E.g., $P_r^{(n)} = \{(Packet\ Throughput)^{(n)}, (Packet\ Latency)^{(n)},\ (Packet\ Loss)^{(n)}\}$
- $T_d^{(n)}$ is the set of delay attributes of state n
    - E.g., $T_d^{(n)} = \{(Delay\ to\ EAS_0)^{(n)}, (Delay\ to\ EAS_1)^{(n)}\}$

The mapping task can be addressed through a quantification of the number of values which the power-specific attributes in the state can take, rather than consider a continuum of values of these attributes. The process of quantification can be guided by the notion of rationalization of a property, i.e., division of its dynamic range into a number of sub-ranges. The granularity of the division will affect the achievement of control objectives and the relationship between the two (granularity and control objectives) is not dealt with further here. Nonetheless, useful rationalization is already available in each component's ***intrinsic*** states, of which the processor core's modes of operation and depths of sleep, are a well-known exemplar. Therefore, proceeding from what is already available in these states, a discretization of the state space can be attempted. Departing from:

- the understanding that it is sufficient to change the value of one property to change state, and
- the expression for G0 illustrating the number of properties pertinent to state,

it is not hard to readily foresee that a general-purpose COTS computer system has a very large number of EASs even in G0. The following consideration elicits this number.

Consider the rth property of processor core C_1. Following point 4 in the preceding numbered list, this is denoted by $P_r^{C_1}$.

Let $P_r^{C_1}$ have $N_r^{C_1}$ possible (rationalized) values.

Furthermore, following point 5, C_1 has R_{C_1} different properties.

∴ Maximum number of possible EASs of core C_1 is:

$$\prod_{r=1}^{R_{C_1}} N_r^{C_1} \tag{4}$$

It easily follows that if there are L cores, then the maximum number of possible EASs of processor cores sub-system is:

$$\prod_{l=1}^{L} \prod_{r=1}^{R_{C_l}} N_r^{C_l} \tag{5}$$

This reasoning can be expanded to cover all M devices in the COTS computer system, as well as the L cores.

∴ Maximum number of possible EASs of a COTS system in G0 is:

$$\prod_{m=1}^{M} \prod_{r=1}^{R_{D_m}} N_r^{D_m} \times \prod_{l=1}^{L} \prod_{r=1}^{R_{C_l}} N_r^{C_l} \tag{6}$$

This number can be reduced to a far smaller one if the maximum number of EASs per COTS computer system (the EAE) is reduced through elimination of its ***unrealizable*** physical states. For example, consider core C_1 again. Suppose that this has $N_{r_{DVFS}}^{C_1}$ DVFS states, where r_{DVFS} refers to the P-states (power-scaling) which C_1 is capable of. Since the processor core's distinguishable EASs are reduceable to these $N_{r_{DVFS}}^{C_1}$ states (clearly, these are rationalized states; each one of the $N_{r_{DVFS}}^{C_1}$ will correspond to an upper and lower limit of power use), then the number of realizable physical states of the processor core subsystem reduces to:

$$\prod_{l=1}^{L} N_{r_{DVFS}}^{C_l} \qquad (7)$$

Now, if ***usefulness*** is a subset of the realizable, then if the notion of such useful states can be extended over all the D_M devices within the COTS computer system, then the number of possible EASs in G0 reduces to:

$$N_{EAS} = \prod_{m=1}^{M} \sum_{i=1}^{I} N_{r_{major(i)}}^{D_m} \times \prod_{l=1}^{L} N_{r_{DVFS}}^{C_l} \qquad (7)$$

where $r_{major(i)}$ is one of the I properties that constitute device D_m's major determinants (or dimensions) of power use. Note that this line of reasoning has thrown up an attribute of what it means to be part of a rationalized set: the useless, or unusable, do not need to concern us. Thus, through rationalization of the power property into a countable set, through well-defined ***operating modes*** of the component, a rationalized generalization of the number of EASs of a system comprising the components can be obtained[20]. These useful states are conceived as “performance states” by the ACPI specification [199, Sec. 3.6], with reference to memory, hard drives and graphics components (as well as LCD panels and audio subsystems).

Indexing can be partially achieved on the basis of the same principle as has been adopted in the GAL standard (reproduced earlier).

1. The active states are sorted in descending order of power consumption, with $n = 0$ identifying the state with highest power consumption and $n = N_{EAS} - 1$ identifying the state with lowest power consumption.
2. Indexing of the states between $n = 0$ and $n = N_{EAS} - 1$ is less tractable, but an approach can be perceived on the basis of sorting of devices in order of maximum power-consumption.

---

[20] Further reductions are immediately visible, albeit not expressible in closed form without further specification of the devices. Any device that depends on a processor to operate cannot do so while all cores are in sleep state.

a. In most computer systems, the processor is the highest maximum power-consumer. Therefore, a first subset of the N_{EAS} active states is obtained from:
    i. all processor cores operating in ACPI P0 state, and
    ii. all the useful combinations of device performance states.

b. 2^{nd} subset: would comprise exactly the same combination of D- and P-states except that core C_1 would be in P1.

c. Further subsets: once all the possible combinations of processor core power states have been exhausted, the device which consumes the highest maximum-power is used to obtain the next subset of states for enumeration.

d. Termination of grouping of EASs into subsets: this process proceeds until the lowest maximum-power device is reached, after which the process of indexing the EASs is complete.

The process only guarantees a monotonic descent of $\{(Maximum\ Power\ Use)^n\}$ with n if the sorting of the components satisfies the following condition.

Let component X precede component X+1 in the sorting.

If the maximum power consumed by X in its least active state is greater than the maximum power consumed by X+1 in its most active state,

then monotonic descent of $\{(Maximum\ Power\ Use)^n\}$ with n is achieved.

6.6 Analysis of GAL's relevance to NGN cores

GALv2's prospect of successful adoption is a debatable issue, but the hinge on which this turns is more likely to be its accretive prowess than its technical excellence, irrespective of how the latter is construed. This claim is best defended on the basis of:

- an inspection of discernable implications in GAL's and GALv2's architecture, on implementation of the control processes;
- an understanding of trends in current and next-generation control architecture, and
- a comparison of the techniques which emerge from the inspection of GAL, GALv2 and control trends.

6.6.1 Implications for control architectures that support GAL and/or GALv2

The GAL architecture comprises a control hierarchy, but the GALv2 architecture, while implicitly acknowledging such an arrangement of control processes, has not committed explicitly to identification of an NCP entity and an LCP entity. In part, this can be explained by the greater depth of control anticipated outside the scope of the node. Indeed, it has been seen that, within the GALv2 + NFV-MANO framework and architecture, control elements are expected to reside on the NFV

Orchestrator, the VNF Manager and the Virtualised Infrastructure Manager (VIM). Of these three control elements, I have suggested that the VIM's control process is of the same scope as a local control policy (LCP), implemented on each NFVI-node hosting VNFs in the GALv2 domain. Here, domain refers to the set comprising the higher-layer controllers and all network nodes with which these controllers are authenticated (as authorized controllers).

GAL's functional architecture declares three main goals, one of which is "to enable LCPs to interact with control-planes and EAEs by means of standard high-level commands (passed through the GSI)" [200, Sec. 4.2]. Moreover, the standard foresees OSPF-TE and RSVP-TE with green extensions as exemplars of network control policies (i.e., processes thereof). Annex B further suggests that the NCP is a "routing protocol extended to consider energy-aware metrics". There is no further, significant, ***explicit*** articulation of a control architecture, but there is an important ***implication*** in the "standard high-level commands" (Fig. 32). Consider the GAL`_Discovery` command. This "permits to retrieve information related to the resource infrastructure", and purports to be the means by which a higher-level control policy process retrieves architectural information from a lower-level control policy process. Such a relationship includes one between an NCP and an LCP, and therefore suggests an unsolicited communication in the direction from a network controller (which would reside in an OSS, at least for legacy networks) to a network device (device is used in the GAL sense here). The same holds true for `GAL_Release`, `GAL_commit` and `GAL_Rollback`. Indeed, the generic descriptor used throughout [200, Sec. 5] is "command", which clearly conveys the designers' rationale. This centralized control paradigm is not scalable; indeed, Annex D attempts to remedy this, by suggesting event-driven ("trigger") control. However, this is only an "informative" annex; not only is the text of the annex rather loose in its recommendations, but it has no mandate for application. Therefore, its value is arbitrary and the annex cannot be considered as a reliable framework for implementations.

GALv2 does not extend the functional architecture; rather, it addresses "the need to adapt the GAL specification … to the NFV environment … [and] the use of GALv2 in the ETSI NFV architectural framework" [202, Sec. 4.1]. Moreover, GALv2 reduces the categories of operation from 6 (six) to 3 (three), dispensing with `GAL_Discovery`, `GAL_commit` and `GAL_Rollback`. Unlike GAL, the generic descriptor for communications, used throughout [202, Sec. 6] is "message". This term is less evocative; a closer inspection is warranted. Consider Table VIII Table IX , Table X and Table XI .

Table VIII MONITORING MESSAGES EXCHANGED ACROSS OS-MA-NFVO

CreatePmJobRequest	OSS/BSS -> NFVO
CreatePmJobResponse	NFVO -> OSS/BSS
Notify	NFVO -> OSS/BSS

Table IX MONITORING MESSAGES EXCHANGED ACROSS OR-VNFM

CreatePmJobRequest	NFVO -> VNFM
CreatePmJobResponse	VNFM -> NFVO
Notify	VNFM -> NFVO

Table X MONITORING MESSAGES EXCHANGED ACROSS OR-VI AND VI-VNFM

CreatePmJobRequest	VNFM -> VIM
CreatePmJobResponse	VIM -> VNFM
Notify	VIM -> VNFM

Table XI MONITORING MESSAGES EXCHANGED ACROSS VE-VNFM

CreatePmJobRequest	EM -> VNFM
CreatePmJobResponse	VNFM -> EM
Notify	VNFM -> EM
Notify	VNFM -> VNF
Notify	VNF -> VNFM
GetIndicatorValueRequest	VNFM -> VNF
GetIndicatorValueResponse	VNF -> VNFM
GetIndicatorValueRequest	VNFM -> EM
GetIndicatorValueResponse	EM -> VNFM

The messages are paired in request – response fashion. Note that the CreatePmJobRequest originates at the OSS/BSS or the network element manager (EM) and filters down to the Virtualized Infrastructure Manager – which, it has been suggested, implements the data plane device's LCP process. These messages may program behaviours in the targeted data plane device: for example, CreatePmJobRequest includes a reportingPeriod parameter that specifies the periodicity of data collection. Furthermore, GALv2 structurally supports event-driven behaviour through its Notify messages. The same observations made for the Monitoring category hold true for the other two categories (Provisioning and Release).

GALv2 thus supports both synchronous (controller expects periodic feedback) and asynchronous (event-driven) control. The latter's support of threshold crossings is essential to real-time adaptation of network behaviour to traffic load and other external conditions (e.g., temperature and power supply). Therefore, GAL implicitly defines a purely controller-driven control architecture, while GALv2's controller delegates feedback delivery to the data-plane device, thereby supporting SDN's closed loop control characteristic.

6.6.2 Trends in current and next-generation control architecture for traffic engineering

Operator survey suggests that MPLS is the dominant data plane switching technology (see Chapter 8); this should not be surprising as it does match anecdotal experience; moreover, provider bridging is a strong contender, the closer to the subscriber the data plane device lies. A central concern in IP networks is augmentation of least-cost-path-based forwarding by path engineering, i.e., computation of paths that take traffic load over path links into account, as well as service-specific criteria like delay and jitter. It is to this context that power use is added as an ingredient of traffic engineering, thus clarifying the often-repeated claim that power must be considered against the backdrop of service KPIs.

The classical method employed in IP/MPLS networks exercises the label-switched path (LSP) construct. An LSP is a path constructed to meet some objective; the most basic objective is simply to reduce local switching time while following the least-cost IP path to a prefix[21]. The LSP is executed in the data plane through inspection of an MPLS header that carries a stack of MPLS labels. At each node participating in MPLS switching (appropriately referred to as a label switching router (LSR)), the topmost label on a received packet's stack is used to consult the label forwarding information base (LFIB) and thereby identify the destination link out which to forward the packet. The LFIB (which is a data plane object) is populated from the label information base (LIB) by the LSR; the LSR operates the label distribution protocol (LDP) to set up a session with a neighbouring (shares a link with) LSR. Over the session, the neighbour sends a datum consisting of {label; IP prefix}; this datum is referred to as a label binding. There may be several neighbours; the LSR will set up an LDP session with each one and receive a label binding for each IP prefix from each neighbour. The LSR selects the label binding that it receives from the neighbour which is the IGP's next hop according to the LSR's IGP-derived routing table; the label thus selected is referred to as the IGP label. Therefore, in this way, operation of the LDP depends upon successful convergence of the IGP.

---

[21] This rationale used to be valid when switching on the basis of IP packet header content was done in software. It no longer holds true now that ASICs and FPGAs are used for this task.

LDP is not the only source of labels; indeed, for traffic engineering, the classical approach uses RSVP[22], extended to carry MPLS labels and traffic engineering data. The node at the edge of the MPLS domain (appropriately, therefore, this is referred to as an ***edge*** router, or label edge router (LER)) takes on a traffic engineering (control) function: calculating a path through the MPLS domain that satisfies constraints which it receives as (input) parameters, e.g., bandwidth requirement, maximum delay, and maximum delay jitter; my interest, of course, is on green[23] paths, i.e., computing a path that meets an objective that saves power. The LER employs two core components in its task: a path computation algorithm (PCALC, a constrained shortest path first (CSPF) algorithm is a typical choice) and a traffic engineering database (TED). The traffic engineering database is a superset of the IGP domain's topology: for each link and for each node in the topology, it stores attributes that describe relevant characteristics (bandwidth, delay, jitter, green characteristics, etc.). The algorithm uses the TED's data to calculate a path that satisfies the constraints. The path is composed of as an explicit route (of IP addresses) through the domain. Subsequently, RSVP-TE (RSVP with traffic engineering extensions) is used on all LSRs on the explicit route to set the path up. Note the implication: the intermediate routers are loaded with the task of maintaining the path's state information.

The principle driving adoption of MPLS label-switching is ***source-based routing***; this is evident for both cases (the IGP label and the TE label) described. In both cases, however, the intermediate LSRs must participate in the control function of traffic engineering during the entire lifecycle of the path. In both cases, moreover, case-specific supplementary control protocols (LDP and RSVP-TE, respectively) must run and add complexity to operations.

An evaluation of the inroads which Segment Routing has made into the role of primary ***control*** technology for Traffic Engineering across the metro area is readily available: simplicity and scalability are the keys [216], [217], [218], [219]. Apart from simplifying operations through dispensation with LDP and RSVP-TE, it also dispenses with path state maintenance (storage space and storage management) on core switching nodes. Furthermore, Segment Routing architecture's concept of segment integrates into the existing MPLS paradigm, while also being embedded deeply within IPv6 (SRv6). While SR-MPLS and SRv6 are not directly comparable, the facility to take steps towards a centralized controlled architecture with an existing data plane switching technology (i.e., MPLS) is a

---

[22] Another, well-known source of labels is Multi-Protocol Border Gateway Protocol (MP-BGP), which distributes the Layer 3 VPN label. This label is stacked below the transport label (e.g., the TE label, or the IGP (least-cost path) label) by the VPN's edge router and is used once the router at the other end of the VPN is reached.

[23] Note that use of the term "green" is made, instead of other more specific terms, to relate to a generalizing concept rather than a mode of deployment. Particular modes might dictate avoidance of unlit nodes, or avoidance of unlit line cards, or avoidance of nodes that support no green capabilities.

third key for successful adoption; meanwhile, SR-MPLS tackles SR's learning curve should the network operator choose to consider migration towards IPv6.

A consideration of how a green protocol can integrate into SR's *modus operandi* will serve well in an analysis of how well GAL and GALv2 are poised for widespread (therefore: successful) adoption. Two options are already available for the role of control protocol on the controller – core switching node interface: these are PCEP and BGP, both of which support Flowspec additions. The controller receives information about green state on nodes through BGP-LS, which is a vehicle for node- and link-state NLRI (network layer reachability information) disseminated throughout a domain by the IGP. Green state might be carried by IS-IS as a TLV record or as a TLV record within an OSPF opaque LSA. This information is then used by the controller to calculate a green path through the domain when flow characteristics match those that uniquely identify green slice customers. Fig. 39 illustrates the high-level components of this architecture.

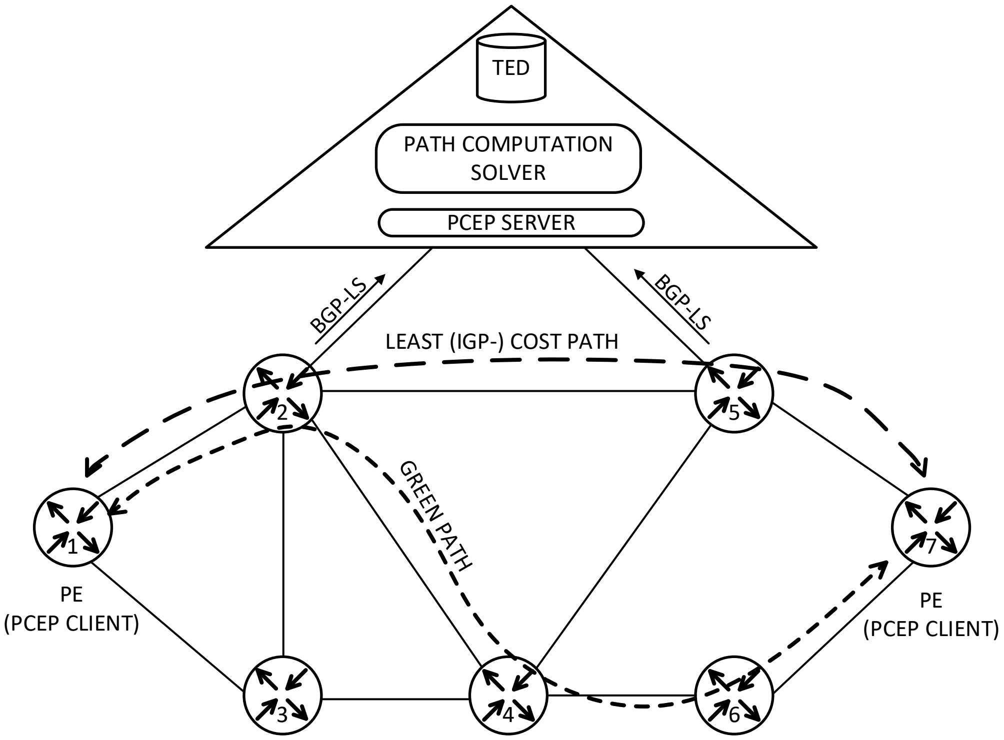


Fig. 39. Green path through a GAL domain

Classical and next-generation control architectures can therefore be contrasted in the following manner.

1. The difference between NGN and classical control architecture can be at least partially expressed in terms of ***the implementational model***. Rather than deploy the (function of

the) path computation element in a data plane network element, it is deployed in a different element of equipment: the SDN controller.

2. Classical control architecture supports an implicit hierarchy, in the form of LERs that are tasked with path computation. The hierarchy is supported through the operation of IGPs that distribute data (as TLV records) on (a) topology, (b) node attributes and (c) link attributes. Both OSPF and IS-IS distribute data by flooding, through adjacencies which routers form over shared links. Changes in topology, node state and link state ***trigger*** flooding (which may be limited in scope). The data is organized by each node into a link-state database and a traffic-engineering database. The LER is a recipient of such updates. Using the updated databases, it can compute paths according to objectives coded in the algorithm.
3. Next-generation control architectures are explicitly hierarchical, with the SDN controller running control functions that can be detached from low latency constraints. Unsurprisingly, the SDN controller employs the same IGPs as the classical approach to control, for the purpose of distribution of data on topology, nodes and links. However, the SDN controller receives link state and other TED data through the means of the exterior gateway protocol: BGP-LS. Thus, the path computation element in the SDN controller can perform the same function as that which is performed in the LER.

6.6.3 The compatibility of GALvX's ramifications with trends in control architecture

As indicated in the previous sub-section, the approach I take is to investigate how well GAL and GALv2 might integrate with an SR control architecture. Within this limited exposition, the purpose is to understand whether the implications for control architectures that support GAL and GALv2, are compatible with that which is already available to implement traffic engineering in a CSP's domain(s).

GAL's control architecture can be characterized as follows.

1. The architecture is a hierarchical one, with an NCP at the peak and several layers of LCP. Its suggested candidate for the role of NCP process is green extensions to extant protocols and/or algorithms; reference to OSPF-TE and RSVP-TE is made.
2. GAL's data distribution is predicated upon a polling mechanism. This applies to all six command categories, and is especially problematic in so far as the `GAL_Monitor` command is concerned (since this is the technique through which green node and link state is collected). This, too, is an unsolicited communication (polling) from the NCP to the LCP.

The following contrasts may therefore be drawn.

1. Classical control architecture is implicitly hierarchical, with the edge devices (the LERs, or the ***provider edge*** (PE)) initiating control transactions. Next generation control architecture is explicitly hierarchical. Therefore, both architectures can be aligned with GAL's hierarchical structure.
2. Distribution of data is based on two events.
    a. One or more nodes is inserted into the subset of the domain affected by the flood scope of the data (i.e., a change in topology has occurred).
    b. A change in node or link attributes takes place.

    There is a limited polling mechanism in OSPF, e.g., during the database exchange process, but this is limited to the scope of an adjacency, and to the purpose of ensuring synchronization of databases [220, Sec. 7.2]. The primary data distribution mechanism is flooding, not polling, and the same can be said for IS-IS.

Summarizing: while the LER (the PE device) bears some resemblance to the GAL entity running the NCP process, there is no structural support for polling in the data distribution techniques prevalent in both classical as well as next-generation cores. This does not mean that a green controller cannot obtain state information from the data plane devices. For example, if the controller supports BGP-LS, then it can obtain topology, link and node data from the data plane's network elements. The unsolicited poll is problematic, however, as there is no suitable vehicle for it in current or next-gen control architectures. As a consequence of the misalignment of control mechanisms, it is unlikely that GAL will be adopted in telecommunications networks.

Analysis of GALv2's control architecture's alignment can proceed in the same way. It has been seen that GALv2's operations are divisible into provisioning, release and monitoring categories. Provisioning and release categories regard operations that are initiated by a controller, and it keeps the initiative throughout their lifecycle. Initiative in monitoring tends to skew asymmetrical towards the data plane device the larger the scope of the data plane grows, under the pressure of the demand of scalability; indeed, this is acknowledged in [200, App. D]. The analysis therefore consists of an inspection of how well the three categories of operation can be accommodated by network control architectures, both classical and next-generation. Consider Fig. 40 [202, Fig. C.1], which shows a superposition of the operations in Table VIII Table IX , Table X and Table XI , over the NFV-MANO architecture for management and control.

Notifications (Notify messages) can be either:

- `PerformanceInformationAvailableNotification`, or
- `ThresholdCrossedNotification`.

The latter aligns well with both OSPF's and IS-IS's modes for distribution of data: ***it relates to change of state***. The performance notification leads to a revisitation of the polling problem. Suppose the notification is in the VIM -> VNFM direction. This leads to a `CreatePmJobRequest`(VNFM -> VIM), which in turn should be followed by a `CreatePmJobResponse`(VIM -> VNFM). Once again, the issue becomes one of perusing the IGPs' mechanisms for one that supports these messages in the course of the mechanism's operations, as an extension. Now, while the `PerformanceInformationAvailableNotification` might perhaps be delivered in the context of a green extension to OSPF or IS-IS, ***there is no detectable means for communicating either*** `CreatePmJobRequest` ***or*** `CreatePmJobResponse`.

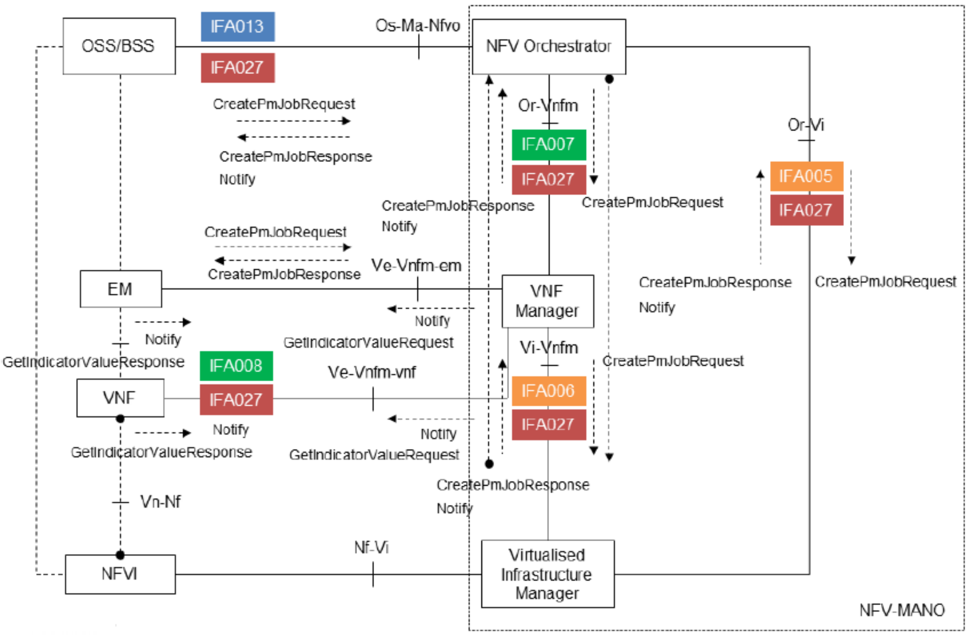


Fig. 40. GALv2 GSI Monitoring operations [202, Fig. C.1]

The prognosis for GALv2's adoption is better than GAL's because GALv2 is nested within a supporting super-structure (i.e., NFV-MANO) that carries enough weight to leverage change. Within any research endeavour in telecommunications, it is well worth keeping in mind that CSPs, notably the larger ones, have vested interests on which returns are sought, and these interests weigh heavily in the balance that is operated to take a decision in favour of, or against, adoption of new technology.

6.7 Case study: power savings through containerized video streamers in an access network

At the time of writing, cloud computing has evolved to the state where a facilitative set of elements has been recognized. The Cloud Native Computing Foundation [221] enumerates, in broad terms, yet with consistency: containers, microservices, service meshes and declarative APIs as elements of ***system infrastructure*** (in a loose sense of the term) that facilitate ***cloud-native computing***. These elements enable software architectures to exploit the scale and flexibility central to the cloud computing paradigm.

Public telecommunications network operators can implement their traditional applications (e.g. operations support systems and business support systems), as well as those more readily associated with over-the-top service providers, in a cloud-native manner. However, concerns with power use of the novel system infrastructure need to be addressed, notably where resource-intensive applications are involved. Video streaming falls into this category, and a case study is presented in Appendix 7. In this appendix, an architecture for a video cache service at the edge of a communications service provider's (CSP) network in the metropolitan area is designed, and a scaled version is implemented in a laboratory environment. The video cache's point of presence (PoP) is representative of an access node/local exchange/central office/distribution hub (AN/LE/CO/DH). The edge network downstream of the video cache PoP uses an optical distribution network (ODN). An intermediate node (IN), representative of a service area node/multi-service access node/subtended access node, is implemented using Open vSwitch (OvS). The access network downstream of the IN is a scaled version of the Active Ethernet (point-to-point topology) type, which was found to be popular in a sample of small CSPs. The architecture's implementation model is described in terms of standardized reference points (RP). The video cache server is implemented using ffmpeg installed as an application of the host operating system (OS) and, separately, as an application in a Docker container on the host OS. Power is measured using both server sensor technology and software (PowerTOP). A comparison is made between power used while streaming videos in both modes (native and containerized) of operation. Containerization is found to incur a low overhead in used power while streaming video, compared with streaming video from ffmpeg running directly on the host operating system. The relationship between the two modes is compared with that found in the literature, notably with respect to conclusions reached in such literature as are obtained with the use of PowerTOP, since limits on the latter's accuracy have been observed.

Chapter 7. Enabling systemic green: predicting power consumption of virtualization containers

The preceding chapter supports a strategy for development of energy-aware performance control in NFV built on the foundations of GALv2, ACPI and NFV-MANO. NFV is couched within SDN's architecture; within SDN there inheres feedback from the data plane and therefore, of measurement of that which it is desired to feedback (or some function thereof). Outside SDN (as well as within SDN), to enable green operations mandates to enable power control; therefore, whether using legacy networks or network function virtualization in an SDN architecture, it is necessary to measure power use. Succinctly, then: ***a general prerequisite to devise energy- and/or power- efficient operations is accuracy in power and energy measurements***. With specific regard to VCs, it is also essential for billing in multi-tenant environments, so that the Infrastructure Provider (IPr) can charge customers the fair amount for the resources (including energy) they consume.

Now, the first question that naturally arises asks whether research into the measurement of VC's energy – and/or power consumption is needed.

- On the one hand, it is fair to claim that precise measurement of a VC's power consumption is difficult, since measurements of its host's power consumption cannot be related directly to it. Hardware power meters are incapable of measuring power consumption of individual VCs co-hosted on a physical machine. Moreover, power consumption of a VC varies with its hosting machine. Therefore, for VCs, accurate measurement is predicated upon precise ***modelling*** of energy- and/or power consumption. It is not a straightforward task. On this reading, research is needed.
- On the other hand, the field of measurement (through modelling) is maturing and the scope for contributing meaningfully is narrowing. This has been shown in a work I co-published, where several ***surveys*** of energy - and/or power models are cited [222, Sec. 2].

However: to be spoiled for choice (of model) is a problem in itself, and while a solution to this problem is the purview of the survey, all the surveys found by us (my co-authors and me) focus on the value (i.e., accuracy, constraints etc.) of the result, i.e., the model. What if a researcher wants to attempt to focus on the narrow scope remaining for research? Knowing what has been tackled within the problem domain, what has worked in tackling the problems, and what range of results has been obtained is of great value to such a researcher. This, then, is the focus of this section: understanding the ***processes*** that inhere to research. I refer to this as the ***dynamics of research*** into a problem domain. Here, the problem domain is modelling the power consumption of virtualised containers in the telco cloud. In this section, therefore, the contribution lies in ***a thorough analysis of the dynamics of research itself: the challenges, the approaches, the pitfalls, the fallacies, the research gaps, without neglecting the fruits***

of research. Dynamics are characterized through a thorough frequency analysis, which is conducted **through application of a novel method** [223], that is unique in its ability to ***parse*** research literature. Through the visual aids provided, and observations through cross-cutting ***themes***, a prospective researcher obtains a thorough characterization of the problems, approaches, developments, formal methods, pitfalls, fallacies and research gaps that characterize this research space

7.1 Understanding the problem domain

The problem domain concerns development of **predictive** energy and power **models**, as well as **measurements** that facilitate qualitative and/or quantitative prediction, of consumption by individual VCs relevant to the telco cloud. A simple interpretation of the rationale that drives this selection is that these works produce results that **measure real-time power consumption by VCs** and/or model **real-time power consumption by VCs**. The object of measurements and modelling is strictly the VC.

Nonetheless, the devil is in the details and so the details of this simple rationale must be worked out. One important, finer point regards the VCs themselves. There are software technologies, which I shall elaborate upon in later sections, which are **functionally critical** to VCs. Works that measure, and/or model, such technologies' power consumption, ***are*** in scope. Justification of this claim on scope is not hard. Since power consumption is a scalar quantity, reduction of power consumption of a component of a VC translates into reduction of power consumption by the VC. In fairness, the translation is not direct (1:1). A generalization of Amdahl's law comes to mind: that improvement in a component, measurable by some metric, is attenuated by the ratio of that component's use (measurable by that metric) to the system's (the VC's) use (measurable by the same metric). Summarizing: research that studies measurement and modelling of power consumption by ***a component*** of a VC is also in scope. For example:

1. **include** the Data Plane Development Kit (DPDK) [224], as it serves the critical function of networking (VCs that serve as virtualized network functions (VNFs));
2. **include** a comparative study that measures power consumption by a VC using two different implementations of input/output virtualization technology, say: SR-IOV (single-root IO virtualization) and paravirtualization;
3. **exclude** a comparative study that measures power consumption by various network adapter (or network interface card (NIC)) architectures, **unless** it reveals the impact of these architectures on VCs' power consumption.

Further detail emerges from the "real-time" requirement. Power meters that emerge from measurements and models thus constrained, may be used regardless of whether instantaneous or statistical readings of power are required. Studies (on measurements and models) meet this requirement by satisfying the following criteria.

1. Predictors:
    a. must be of fine temporal granularity;
    b. must be updated with the regularity of the temporal granularity, and
    c. must enable prediction of power consumption at the same temporal granularity.
2. Workload: Only resource-specific constraints are considered. That is: in the course of testing using, say, workloads that are processor-intensive (hence the workload is specific to the processing resource), no other constraints are allowed in works included in this study. In particular, models must not constrain the stochasticity of the workload.

The second detail regards workload. It is useful to anticipate that the universal power model is a fallacy and the principal reason is that the relationship between workload and architecture cannot be pinned down indefinitely. This does not mean that modelling is a fruitless endeavour. It simply means that validity constraints must be placed on the model in terms of workload and architecture. Therefore, modelling and measurement is not excluded because of its workload-scope or architectural scope. It is important to observe, early on, that power models of VCs are pitfalls for those who apply them without knowledge of such limits

7.2 Organization of this chapter and an outline of some major observations

This chapter is organized as follows.

1. Appendix 2[24] [223] describes the method, which produces a number of problem-approach-development (PAD) triads that, when aggregated, produce statistics that characterize research dynamics. To the best of my knowledge, this is the first time that this method has been used. Description of the method has been delegated to an appendix, to resolve the difficulty of elaborating fully on the method without distracting attention from the results which have been obtained.
2. Section 7.3 presents the detailed results.
3. Section 7.4 presents the analysis of the results. There, a qualitative assessment is given through themes which emerged as the data was organized. These themes have been classified as “state of the art”, “fallacies” and “pitfalls”, to suggest guidance and warnings which were gleaned from others’ experiences.

---

[24] https://www.sciencedirect.com/science/article/pii/S2215016122000188

4. Section 7.5 concludes by attempting to encapsulate the insights gained through this work.

This chapter is complemented with an online repository[25], that carries the raw data.

Among the themes that this survey identifies: all problem categories found touch one or more of a set of seven main variables that may affect power consumption by virtual entities and the ensuing model representations: workload type, characteristics of the virtualization agent (VM or container), host machine resources and architecture, temperature, operating frequency, attribution of a fraction of consumed power to individual VCs, and mutual influence of concurrent VCs.

Among the major pitfalls that emerged from the thematic analysis, I highlight here the misconception on the Data Plane Development Kit's (DPDK) power efficiency (commonly misportrayed as a power hog), the often-unacknowledged limitations of the widely used linear models, problematic use of benchmarks in model validation, the failure to precisely identify the physical context in some experimental research, the influence of synthetic workload generators on measurements, and the sometimes-overlooked relevance of processor organization on power consumption measurements. I have also pointed out the unavoidable need to precisely identify the scope and limitations of models and the fallacy of the quest for a "universal" power model.

The research gaps identified are four:

1. modelling of containers' power consumption;
2. the effect of overcommitment on power efficiency;
3. investigation and classification of DPDK applications, and
4. modelling of power consumption by virtualized I/O (a challenge which is starting to receive some attention).

7.3 Survey results: a digest of challenges, approaches and developments

7.3.1 A taxonomy of the problem space

As parsing of the literature proceeded (method described in [223]), it was observed that the scope of this survey is relatively narrow and the problems in the set are not fully independent of one another. Rather, the problems diverge from one another only as ***aspects*** (it could also be said that they are ***derivatives***) ***of the core challenge of modelling the power consumption of virtualized containers***. Each RU (research unit)[26] is rooted in this core challenge, but the derivative problems (the *Problem*

---

16. https://github.com/humaira-salam/PowerMeasurementAndModelingRawData
17. The RU, or unit of research, is "a publication (excluding surveys) in conference proceedings and journals" that "ha[s] three common manifest properties", i.e. problem(s), approach(es) and development(s) [223].

categories (P-categories) and their members) addressed differ from one RU to another. Fig. 41 is an illustration of a simple organization of the challenges which have been tackled in the literature and shows their frequency of presence in RUs. The organization gives prominence to how challenges have been perceived:

1. One group regards the concern with obtaining an understanding of the dependency of power consumption on some genre of artifacts. Categories P1,2 and P9 – 11 are in this group.
2. The other group regards the concern with how to predict power consumption. Categories P3 – 8 and P12 are in this group.

I now proceed to describe the categories in more detail. Each description is preceded by a list of references to works that tackle a challenge in the problem category.

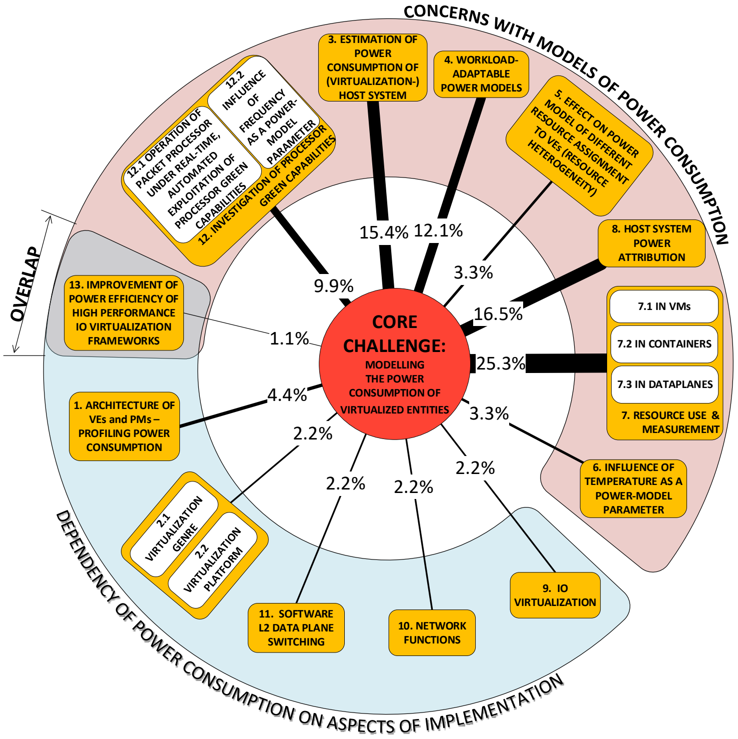


Fig. 41. The core challenge and its derivatives; research interest R_(P_k)is shown in line thickness and as percentage

1) Problem category P1: Host system hardware architecture perspective: dependency of VC power consumption on host system architecture [225], [226], [227], [228].

Challenges in problem category P1 address the impact of specific architectural attributes of the host system on power consumption of VCs. They relate to changes in power consumption (the behaviour) as major attributes of architecture and system-level design are adjusted, inserted or removed.

Insertions and removals are coarse configurative actions such as enabling or disabling; adjustments consist of progressive modifications such as adding increments of a resource. Examples of attributes which have been tackled include multiple processor cores, processor frequency scaling, Non-Uniform Memory Access (NUMA) and hardware threads (e.g., Intel Hyper-Threading). For network functions, the importance of knowledge about the power efficiency of NUMA and multiple-core architectures has the added relevance of these architectures' relationship to ***determinism*** [85, N. see video @19:00, @25:25]. I dwell further on the underlying premise of ***hard partitioning*** during consideration of the impact of the high-performance data plane on power efficiency (see sub-sub-section on DP''K's power efficiency).

Research that investigates the dependency of power consumption on architecture is ***exploratory, charting*** work. It attempts to provide a framework for detailed modelling through the discovery of broad relationships. Problems in this category arise with developments in architecture and system-level design. For example, while [226] is a comparatively old work that tackles architecture, [227] is newer and finds scope for research in system software's exploitation of NUMA. One recent, highly significant scope is that of the use of ***domain-specific architectures (DSAs)***. Researchers are exploring specialized hardware in the quest for improvement of the energy-performance-cost ratio, and will investigate energy efficiency in the process of their research. As domain-specific cores are mixed with general-purpose cores, many architectures will be investigated from each of the three pinnings: energy consumption, performance and cost. A particularly relevant set of DSAs regards real-time packet processing by computer systems hosting NFs at intermediate nodes (INs) at the network edge. Concern lies with expediting the common tasks, such as sending/receiving packets and processing headers. SR-IOV is a good example (see, for example [229], and its inclusion in [225]), but software-only solutions, such as poll-mode drivers, may also help to cut through the many middlemen characteristic of general-purpose operating systems [225], [230]. Introduced to serve the perspective of performance, it is now necessary to understand their impact on power efficiency. Therefrom, it is necessary to understand how to ***control*** their power consumption. I suggest that "***profiling***", the term chosen in [226], [231], is a helpful descriptor of this kind of research. Just as a profile produces an external boundary within which to fill detail, so does this kind of research provide a framework through which modelling work is facilitated and Iin whic157odellingng work provides details of power consumption relationships.

2) P2: impact of alternative virtualization genres and virtualization platforms on VC power consumption [232], [233].

Here, a behavior that consumes power is investigated across different ***implementations*** of a ***system concept***. Investigation of two different system concepts has been observed: (a) virtualization genres and (b) virtualization platforms. There are three members of the virtualization genre group:

containers, para-virtualization and hardware-assisted virtualization. In the virtualization platform group, examples include Xen, Hyper-V, Kernel Virtual Machine (KVM), Docker and Linux Containers (LXC). Research questions typical of category P2 seek to control scope of experimentation through exercise of specific resources, e.g., per-host networking using emulated switches (software switches) [232], [233], processor-bound and memory-intensive processing [54].

I consider genres and platforms as sub-categories of the same overarching problem category. Namely, this is system-level exploration that attempts to establish generalizations about an uncharted space. Like problems in category P1, new problems in this category arise with fresh alternative virtualization genres and platforms. However, here the scope of investigation is broader than with works classified under P1. Unlike P1, where specific architectural aspects (e.g., NUMA, hardware threads) are explored, the perspective taken here is concern with the impact of a choice of implementation of a system.

3) P3: Estimation of power consumption of (virtualization-) host system [234], [235], [236], [237], [238], [239], [240], [241], [242], [243], [244], [245], [246].

Measurement of a single server's power consumption through the use of an external power meter is a trivial task. However, at the scale of cloud datacenters, it is a logistical burden. In addition, travel to the datacenter's site may be burdensome. Furthermore, service availability would be reduced by the process of attaching a physical power meter to the hardware in the virtualization platform, e.g., between the server's power inlet and the outlet in the racking cabinet's power distribution unit (PDU) (naturally, availability would only be affected in cases that do not integrate (management and) measurement facilities within the PDU).

The alternative is deployment of software power meters. In the scope of this survey, the cases considered are meters that attempt to predict host power consumption on ***the basis of activity in the VC***. This challenge is tackled, for example, in [235], [236], [237], [239], [244], [247]. These works then proceed to tackle the problem of attribution of system power to the guest VCs. Indeed, inclusion within scope of both challenges (modelling power consumption of VCs and that of the host system power) seems to significantly enhance the usefulness of such research, with relatively less effort.

Host power consumption may be predicted in terms of VC resource utilization, or in terms of ***a simple characterization of the VEs' workload***. Use of simple workload characterization as predictor requires knowledge of workload parameters like number of processes, number of threads, web interactions per second and network interface utilization. Enokido's and Takizawa's work [244], [245], [248] is noteworthy in its consistency in modelling in these terms, but other variants of this approach have been found: (a) web interactions per second [242] and (b) number of VMs running processor- and/or network-intensive workloads [243].

To contrast: works like [234], [249], [250], [251], [252], [253] are not included within this category – notwithstanding their development of models for prediction of host system power consumption. In these works, host system models were developed as part of the scope of the challenge of modelling virtualized entities. Therefrom, the challenge of system power attribution (problem category P8) was tackled to proceed to guests' power models.

4) P4: Dependency of power model on workload [233], [239], [244], [245], [246], [248], [252], [253], [254], [255], [256].

This category regards the perceived dependency of a VC's power consumption model on the tasks it is processing. While it is intuitive to expect power consumption to depend on workload, it seems far less intuitive to expect the model to depend on workload. If this dependency is detected, the problem of model formation must undertake this aspect of investigation. Two different, major approaches towards achieving ***adaptability*** of model to workload have been observed.

1. ***Adaptation during run-time:*** the selected mode of instrumentation may not be suited to a generalizable, closed-form relationship between inputs and power consumption. In this case, model parameters must be re-trained online. This approach is therefore of the operating-time, or run-time, kind.
2. ***Off-line adaptation:*** a larger set of inputs may need to be identified to comprehensively characterize variation of power consumption with workload. This approach is therefore of the design-time, or off-line, kind.

I conclude this part with a note about two descriptors of workload: homogeneous and specific. The term "homogeneous" is encountered in literature to refer to the case where host system deployments within scope are subjected to a single workload. The term seems to originate in warehouse scale computing (WSC). Conclusions drawn from this kind of workload have drawn criticism as the results, while significant by virtue of the mass of WSC, are not generalizable. The other term – "specific" – identifies a single application; for example, a member of the Standard Performance Evaluation Corporation (SPEC) CPU2006 suite [257]. This term is used to indicate that models tested under such a workload are application-dependent and are valid only within a limited range of this dimension of variability (i.e., the "workload" dimension, see the treatment of the seven dimensions of variability).

5) P5: Dependency of VC's power consumption and power model on VC's resource configuration (heterogeneity) [239], [250], [255].

This category regards the perceived dependency of a VC's power consumption and/or the dependency of its power consumption model on (a) the physical host's resource configuration and (b) the individual VC's resource allocation. Research here is concerned with two cases of very practical problem: the impact on power consumption of (a) the differences between hosting machines/containers and (b) the differences between virtual machines. I have observed that occurrences of research that undertake the challenge tackle it as an adjoint to another focus, not as the research’s primary objective.

1. ***Physical host configuration:*** Host machines in a cloud datacenter may be expected to come in a limited variety of types, principally differing in resource capacities such as the number of processor packages per server, cores per processor package, amount of RAM per server, spread in storage device sizes, etc. Processor power consumption is notably variable, even within a single family of processors. Indeed, specialization in optimized power consumption within a family of processors is part of the study carried out in [82] within the context of an edge cluster for use in NFV. As a VC migrates from one processor within a family to a processor of different specialization, its power model will change.
2. ***Individual VC’s resource allocation:*** The power consumed by a VC varies with allocation of resources to (i.e., in use by) a VC, which can be dynamically varied. The number of virtual cores assigned to a VC is a notable example; see, e.g., [234], [238], [249], [250], [251]. Moreover, VCs are commonly offered in sizes, e.g., small, medium and large, where allocation varies within all the major resource categories, demanding prediction of power consumption matched to the size of the purchased VC.

6) P6: Impact of temperature and/or frequency on models that p’edict V’s' power consumption [237], [258], [259].

This category regards the challenge of inclusion of processor package temperature in models of power consumption. Works that tackle this challenge are concerned with detailed models of power consumption. Here, interest lies in obtaining models that incorporate dependence on hyper-parametric attributes like temperature.

7) P7: Loading the VC’s resources and measuring resource use [233], [234], [235], [236], [238, p. 20], [239], [242, p.], [245], [246], [248], [250], [252], [253], [254], [255], [256], [258], [259], [260], [261], [262], [263], [264].

This category regards use of computing resources and measurement of such use by VCs. Interest stems from the role of resources ***as predictors in modelling***. The researcher is firstly concerned with ***loading*** (i.e., effecting use of) resources. What means ***within the operating context of the VC*** can be used to load a resource? Should it be loaded in isolation (using synthetic loads) or should it be loaded using representative (realistic) workloads? Once these problems have been addressed, the concern with ***measurement*** of resource use arises. The problem here consists of identifying the means that quantify resource use made by the loading.

8) P8: Attribution of host system power consumption to individual VCs [234], [235], [236], [238], [239], [250], [252], [253], [254], [256], [258], [259], [261], [263].

Attribution of host system power to individual VCs is a fundamental problem in proceeding from the directly measurable (host system power consumption) to the indirectly measurable (individual VCs' power consumption). Direct measurement of host system power is possible (e.g., at the wall outlet, or through voltage rail in-line resistors), and such empirical evidence can be used as ground truth and compared with power consumptio161odellirred throug161odellingng. How, then (and herein lies the problem), can this consumption be attributed t' the ho't's individual guests (the VCs)?

Within the host system, power consumption may be divided into ***idle (static), active (dynamic)*** and ***overhead***.

1. Idle (static) power consumption
    a. Power consumption while idle is not attributable to any VC at all, as this consumption arises out of electronic behavior of semiconductor material, not of computation, communication or storage.
    b. Nonetheless, this power consumption must be accounted for and different approaches have been followed. For example: the physical machine's idle power is attributed to individual VCs in fractions equal to the ratio of each VC's virtual CPUs (vCPUs) count to the total complement of vCPUs active on the physical machine [234], [249], [250], [251]
2. Active (dynamic) power consumption
    a. The active component can be linked to a particular VC.
    b. This includes active power consumption in peripherals, e.g., network interface cards/adapters (NICs) and mass storage devices.
3. Overhead, e.g.:
    a. operation of heat dissipating units (fans) to prevent thermal runaway, and
    b. losses in the power supply.

A "top" (host) – "down" (guest) approach to attribution has been observed.

1. Decide on what host system power consumption is within the scope of the study and how to divide it. The problem of attribution of the above three causes may be summarized as follows:
    a. Is idle power attributed to the VCs or is it attributed to the host/a privileged guest?
    b. Is consumption by peripherals within scope? How will this be attributed?
    c. Are overheads modelled or is correlation with other sources of power consumption going to account for them?

2. Select a set of performance metrics that are correlated to a VC’s power consumption.
3. Select a model that maps a VC’s performance metrics to its power consumption.
4. This fourth consideration is tackled only by those researchers who investigate the ***adaptability*** of the attribution obtained through steps 1,2 and 3. Does the obtained attribution adapt well to concurrent, co-hosted VCs? That is: if concurrent, co-hosted VCs were to be investigated, would the division, metrics and model still result in accurate prediction?

9-11) P9: Implementation of virtual I/O; P10: Implementation of network functions; P11: Implementation of software Layer 2 (L2) data plane switching [225], [228], [230]

These three categories are introduced together, since elements from the respective categories are commonly implemented ***as a set*** for the purpose of realization of virtualization of network functions. Here, researchers seek comparative statements and/or broad correlations (e.g., independent, positive, negative) between workload (often in terms of packet rate and size) and power consumption, across implementations of the same type. As was observed for categories P1 and P2, researchers seek a profile of the power characteristics of implementations. It may be helpful to repeat that by “profile”, I understand that characteristics sought here are not closed-form expressions. Examples of elements from the respective categories are:

1. Virtual I/O (P10): virtio [265] and DPDK poll-mode drivers (PMDs) [266]
2. Network functions (P11): Bro (now Zeek) [267] and Snort [268]
3. Software layer 2 data plane switching (P12): Open Virtual Switch (OvS) [269] and VALE [270]

Problems in each of these three categories merit separate classification as they have been tackled separately in the literature. For example, in [230], a number of components are investigated: three different implementations of software virtual switch (P12), two different I/O virtualization devices and two different implementations of the same network function (intrusion detection system (IDS)). In [228], power consumption by packet transmission under DPDK is investigated under the condition of enforcement of (a) the network adapter's affinity to NUMA nodes and (b) DPDK process pinning to processor cores.

12) P12: Investigation of processor green capabilities [227], [237], [244], [258], [259], [260], [262], [264], [271]

Works in this category investigate low-power-idle (LPI) and adaptive rate (AR) operation of a processor as a means of reducing power consumption. The challenge is broad enough to permit a sub-categorization into (a) those works that investigate the influence of frequency as a power-model parameter [258], [259] and (b) other works that address improved, real-time governance of LPI and/or AR [260], [264], [271] to minimize the power consumed to process a load.

13) P13: Improvement of power efficiency of high-performance IO virtualization frameworks [264]

A separate classification was set up for [264] as this work represents an evolution of those classified under P9. This work extends beyond profiling and suggests use of low-power-instructions as the means to balance performance and power efficiency.

7.3.2 A taxonomy of approaches

Fig. 44 illustrates the taxonomy I use to structure the approaches detected in research work. Line thickness and percentage values represent the ***utility*** of the specific approach. Utility is best understood within the context of all the observed triads in a literature corpus. For any specific approach, this may be used to solve a variety of problems and its application may result in a variety of developments. One may therefore think of the approach as a ***nexus***, or a point of confluence through which many researchers pass as they attempt to solve problems. Thereafter, researchers diverge radially outward from this point of confluence towards some achievement (some development). A suggestive image is to think of the approach as the centre of a star, but spokes converge onto it from problems and diverge away from it onto developments. When the count of these triads (each composed of two radial lines, or dyads) is divided by the sum of all such counts for all approaches, a metric is obtained: a normalized quantity obtained within the context of all approaches. A formal statement of ***utility*** may be found in the description of the method in Appendix 2[27].

I now proceed to describe the categories within the context of the taxonomy. Each approach-category's description is preceded by a list of references to works, each one of which uses a component within that category's set.

---

[27] https://www.sciencedirect.com/science/article/pii/S2215016122000188

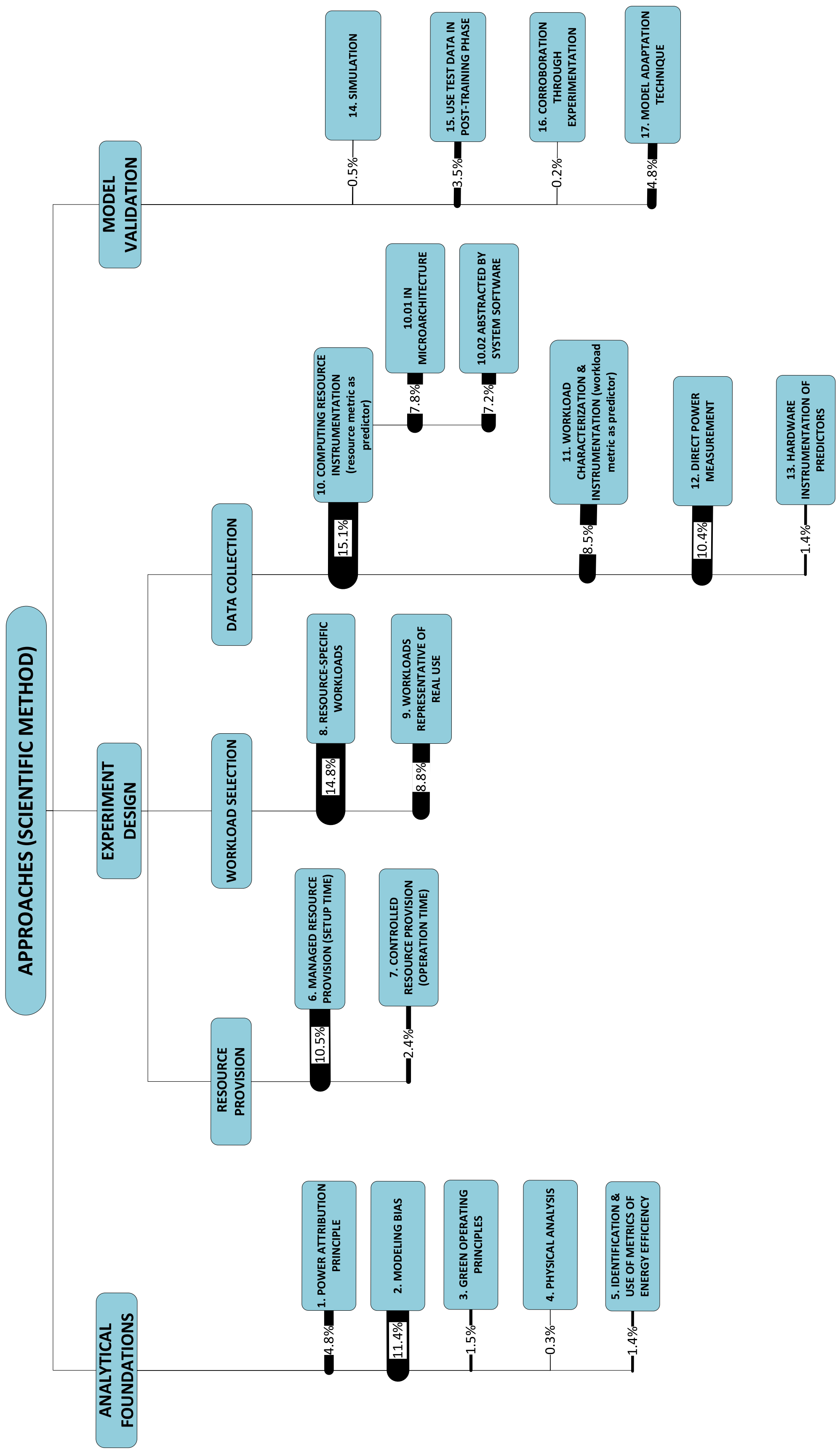


Fig. 42. A taxonomy of approaches; approach utility U_{A_k} is shown in line thickness and as percentage

1) Analytical foundations

This group of categories regards the theory and hypotheses that comprise the essential abstractions at the basis of scientific research.

A1: Power attribution principle [225], [230], [234], [235], [236], [238], [239], [246], [250], [252], [253], [254], [256], [258], [259], [261], [263]: When *host system* power is measured, whether at the wall outlet or at one or more of the power supply's output lines, there is the problem of attributing the measurement to the logical divisions (VCs) of the host computer system. Attribution of system power to modelled entities starts with a decision on which power consumption is within scope (see section 2.1.8). Next, power in scope is attributed to (burdened on) one or more entity. For example: will idle power consumption be attributed to the host system or will it be attributed to the VCs?

A2: Modelling bias [234], [235], [236], [237], [238], [239], [240], [241], [249], [252], [253], [254], [256], [258], [260], [261], [262], [264], [271]**:** Researchers approach the problem of developing a model under some bias which conditions their final outcome. This bias is manifest in researchers' selection of a particular type of regression to apply to their data. Note that this same observation is carried in [272]. Here, my purpose is solely to draw attention to what has been observed as researchers' ***modus operandi*** without analyzing their choice of approach.

A3: Green operating principles [260], [264], [271]**:** Works in this category weave radical approaches to power efficiency into their developments. For example, instead of conventional scheduling, ***run-to-completion*** [273] is exploited to obtain dedicated (or, at least, very sparsely shared) resources for the processing of packets. This approach is further nuanced by real-time control of adaptive rates and sleep depth. In one particular case [264], the novel concept of a low-power instruction instead of transitions to/from low-power idle (sleep) states is used.

A4: Physical analysis [237]**:** This category regards approaches rooted in physical properties of (semiconductor) material in the consumption of energy. Only one work [237] was found fitting this category. However, another two that used this approach to study power consumption of physical entities (and therefore outside the scope of this study, which is concerned with VCs) were found and they are described next to illustrate the approach better. In [247], a study implicitly applies Dennard's law in the process of obtaining weights that scale a processor sub-unit's contribution to power consumption. In [274], the physical cause of power consumption in metal-oxide-semiconductor (MOS) material is examined and used as the basis fo165odellingng equations.

A5: Identification and use of metrics of energy efficiency [226], [227], [242], [262]**:** The relationship between system architecture and power consumption can be investigated through identification of relevant metrics of energy efficiency. For example, an easily recognizable metric, albeit rather broad in possible interpretations, is the J/b (joule/bit). Use of such metrics encourages joint consideration of function and power consumption.

2) Experiment design

The practical, hands-on aspects of the empirical process are the product of a (probably cyclical) design phase, concerned with a number of stewarding activities pertinent to test subjects and ancillary objects in the testing scenario, instrumentation, inputs and outputs. The activities include selection, configuration, interconnection, initiation, observation and termination. I have identified several examples of such activities within this part of my research's scope and grouped them under ***resource provision*** (categories A6, A7), ***workload selection*** (A– - A9) and ***data collection*** (A1– - A13). I describe these categories next. Admittedly, the activities referred to (i.e., selection, configuration, etc.) have broad meaning; therefore, in the course of describing the categories, references to the activities are emphasized by bold, italicized text.

A6: Managed resource provision [225], [226], [227], [228], [230], [232], [233], [237], [243], [244], [245], [250], [258], [259], [260]**(*selection*, *configuration*):** This concerns the provision of resource capacity either to a specific VC (the guest system) or to the physical entity (the host system) hosting the VCs. Within the empirical process, the techniques in this category provide the means to observe the effect on power consumption of managed changes in resource provision. Examples include (manual) pre-***configuration*** of:

- frequency of operation of processor cores [225], [227], [258]
- core affinity [228], [232] and hardware-thread affinity [245]
- network interface data rate capacity capping [243]

These techniques are executed as part of the process of ***selection*** of operating parameters, i.e., setting up experimentation, ***before*** operations start. This qualification is necessary to distinguish from such approaches as may change the operating, run-time context.

A7: Controlled resource provision [255], [260], [262], [264] ***(configuration)*:** Provision of resources may change during the running of an experiment (rather than before it starts). Approaches in this category include the automated adjustment of:

- processor frequency (also known as performance state, or P-state) [255], [260], [262], [264]
- depth of processor sleep (also known as low-power-idle-state, or C-state) [260]
- number of hardware threads [255]
- time spent running a low-power instruction [264]

These techniques are approaches to solving the problem of full-throttle operation. Without guided operation of adjustments like those listed above, operation of the processor may quite reasonably be likened to a multi-assembly-line manufacturing plant that operates line machinery whether there are goods to produce or not.

A8: Resource-specific workload, A9: Representative workloads *(selection, configuration, interconnection)*: These two categories regard the *workload selection* stage, within experiment design in the scientific method. The workload comprises the *inputs* referred to earlier; inputs must be *interconnected* to the system under test, and this is often not a trivial task. In the thematic analysis in Section 7.4, I identify *workload type* as one of the seven dimensions of variability of power models. The influence of workload type on the model obtained is evident in the attention paid by researchers to selection of workload type. Two broad categories of workload type are distinguishable.

Resource-specific workloads (A8) [225], [226], [227], [228], [230], [232], [233], [236], [237], [238], [240], [241], [243], [244], [246], [248], [252], [253], [254], [261], [263] are applied to investigate the impact of utilization of specific resources on power consumption. Such ***synthetic*** workloads are applied ***(interconnected)*** to a machine (whether virtual or physical) to reduce (as much as possible) the scope of power-consuming resources to a targeted set. Resource-specific workloads are most commonly used in exploratory work, to gain an understanding of the behavior of a resource's power consumption. I refer to this approach as resource-specific workloads, or synthetic workloads, or resource isolation.

Representative workloads (A9) [225], [233], [234], [235], [237], [239], [242], [245], [246], [250], [252], [255], [256], [258], [259], [260], [261], [262], [264] may be used complementary with, or alternative to, resource-specific (synthetic) workloads. A notable complementary use is made in the testing (post-training) phase of model development, when representative workloads are used to validate a model (obtained using synthetic workloads). They may also be used in an entirely alternative approach to synthetic loading, to support development of application-agnostic models. Representative workloads lead to training data that incorporates variation in utilization of more than one resource at a time; hence providing at least limited application agnosticism.

The next four categories (A10 – A13) regard the ***data collection*** stage, within experiment design in the scientific method. Categories A8 and A9 regard ***selection*** of workload type. Approaches described here regard measuring ***how much*** of a resource is being used, or workload is being applied, and how long to apply the workload to obtain statistically valid results ***(initiation and termination)*** Categories A10, A11 and A13 regard instrumentation ***(observation)*** of those variables considered (by the researchers concerned) to be reliable predictors of power consumption.

Resource instrumentation in microarchitecture and system software (A10) [225], [226], [233], [235], [236], [238], [239], [240], [241], [246], [250], [252], [253], [254], [255, p.], [256], [258], [259], [260], [261], [262], [263], [264] includes approaches that measure resource use. These measurements are then used to predict power consumption. I make a somewhat weak distinction, for reasons I shall refer to, between instrumentation of microarchitecture and instrumentation by system software. The former regards parts of the processor interface that address the processor's infrastructure for monitoring: event counters (e.g., instructions retired, last-level-cache (LLC) misses, translation lookaside buffer misses) and, more recently power counters (e.g., Intel's Running Average Power Limit (RAPL) [275]. On the other hand, system software's instrumentation is carried out through intermediary system software and includes, most notably, processor utilization. I have seen references to these two categories as "hw counters" and "sw counters" respectively [276]. The distinction is weak since system software is increasingly exposing microarchitecture instrumentation (consider, for exam'le, Lin'x's *perf* tool). This reduces the need to directly access hardware registers and blurs the separation between what is abstracted and what is concrete, raw, hardware data. Nonetheless, through understanding of the data used, the approaches have been separated into two sub-categories.

Category A11 regards use of a simple characterization of workload as predictor of power consumption [225], [226], [230], [241], [242], [244], [245], [248], [255], [260], [262], [264]. This is notably different from approaches in category A10, which are concerned with resource use as predictor of power consumption. Other examples (apart from those given earlier) of workload metrics as inputs are "number of processes" [248] (from the same process image), "transmission rate" [245] and millions of instructions per second (MIPS) [241].

Characterizations may need to be sharper. For example, since packet network traffic arrival is known to often have the properties of a Batch Markov Arrival (stochastic) Process (BMAP), this is an operating constraint (***selection, configuration*** and ***interconnection*** are all ingredient activities here) adopted in the approach of several works studying power efficiency of network functions [260], [264], [271], [277].

Category A12 refers to the measurement of (host) power consumption [225], [226], [227], [230], [232], [233], [234], [235], [236], [238], [239], [240], [241], [244], [245], [248], [250], [253], [256], [261] , ***(observation)*** which is most usually measured at the wall outlet or at the power supply inlet. More granular approaches are desirable, and indeed cases can be found [239] that attempt measurement at the power supply output. The principal drawback of such granular techniques is not (principally, at least) construction of intermediary hardware (e.g., riser boards, or line resistors) but the difficulty in attributing power drawn through any single dc voltage output (or group thereof) to particular consumers. With the advent of RAPL and certain guarantees on its accuracy, the need for direct power measurement has been, at least partially, avoided.

Category A13 regards use of hardware sensors to obtain inputs and/or parameters for the power model [237], [258], [259], [264]. These include:

1. voltage sensors (processor supply voltage)
2. temperature sensors (processor package and memory temperatures)
3. fan speed sensors (processor and chassis fans)
4. wall-clock time measurement

Some of these variables are used in models that predict power consumption while accounting for the effect of drift of temperature and automated supply voltage adjustment (in dynamic voltage and frequency scaling – DVFS).

3) Model validation

Model validation is a multi-faceted endeavor and this is reflected in the approaches that have been detected. The approaches range across the candidates that would typically be considered: simulation (A14) [264], use of test data (A15) [234], [239], [240], [241], [254], [262] and corroboration through experimentation (A16) [260], [264]. I skip elaborating on these categories as they are either self-evident (A15) or because they are too rarely used to permit general commentary. However, to these three categories I add a fourth (A17), namely: model adaptation technique, which I describe below and explain why it fits within this branch of the taxonomy.

Model adaptation technique (A17) [234], [235], [236], [238], [239], [244], [245], [246], [248], [250], [252], [253], [255], [256], [258], [259], [261], [263] : This refers to the approach(es) taken (if any) to develop an adaptable model or modelling system. Here, ***adaptability*** refers to the fitness for use which the model exhibits under change in one or more of the seven dimensions of variability that will be defined in detail in sub-section 3.1.1 below. Model adaptability is essential for practical virtualization, where changes in, say, the number of co-hosted, concurrent VCs, or in workload type are commonplace. Here, I list the major approaches taken towards producing adaptable models. Since these approaches emerge in the context of validating a model's accuracy under some limited range of the seven-dimensional space of operating conditions, this category of approaches is classified as an aspect of model validation.

1. ***Adaptation to change in the number of co-hosted, concurrent VCs*** is widely achieved through ***time-division multiplexing*** of event counters [236], [239], [246], [256], [263], RAPL counters and CPU utilization [238], [250]. This approach enables use of such metrics as predictors of dynamic (active) power consumption, by apportioning counts to VCs in accordance with the time during which the VCs were active.
2. ***Adaptation to uncorrelated causes of power consumption*** can be achieved through ***additional predictors*** [259], to follow causes of power consumption that do not correlate well with counters within the current set. This case reflects itself as poor accuracy in predicted power consumption. Although counter-based models are reported to fit a variety of processor- and memory-intensive workloads well, it may be necessary to account for unanticipated activity through the approach of adding previously unused counters.
3. ***It may not be possible to fit a single model with parameters known a priori***, to the whole range of inputs within the scope of study, notwithstanding the diversity of predictors employed in this pursuit. The following adaptive techniques have been found in the literature.
    a. ***Dataset partitioning***, where the dataset is of the form { {predictors}, response } is used in [254] to match the best model out of a set of models to current, actual operation. An individual model in the set is associated with a single node in a decision tree and the node is selected according to features pertinent to current, actual operation. A simpler, but conceptually similar, approach is taken in [235]. A number of models are devised and model-selection features are limited to the number of active VMs and a coarse grading of CPU utilization.
    b. ***Modelling on demand*** is the term I use to succinctly refer to the fourth adaptation class of techniques.
        i. One early example of this approach is found in [252], where the dependence of the model on workload has been addressed through ***online training***, whenever prediction accuracy of extant models falls out of a range of tolerance. The rationale adopted is that if model

adaptation to such an "unseen" case is limited to parametric tuning, then a modelling system might be able to construct a model while VCs are in operation.

ii. A broader perspective is found in [255]. An automated system for profiling containerized applications is described and demonstrated. Containerized applications are profiled from three perspectives: computing resources consumed, energy consumed and performance. In this case, the rationale is that energy consumption can be optimized by determination of a frequency-and-hardware-threads host configuration that meets performance requirements. Thus: starting from central functional requirements (performance requirements), operating conditions are determined that minimize energy consumption. This approach is capable of meeting the challenges posed by heterogeneous host hardware and application (workload) diversity, at the cost of analytical modelling. Indeed, characteristic curves can be derived but causes underlying observed behaviours remain unaddressed.

7.3.3 A taxonomy of developments

Developments fall cleanly into one of two groups: (a) models of power consumption and (b) observations on dependencies of power consumption. The first group (D1-D10) includes developments that ***predict power consumption*** over a sub-space of the seven-dimensional space of operating conditions. The second group (D1 – D18) includes developments that are oriented towards the ***correlation of power consumption*** with aspects of system integration. As the taxonomy is rather broad, it is presented here in three parts:

1. Fig. 43: top-level fork into models and dependencies;
2. Fig. 44: the taxonomy of models, and
3. Fig. 45: the taxonomy of dependencies

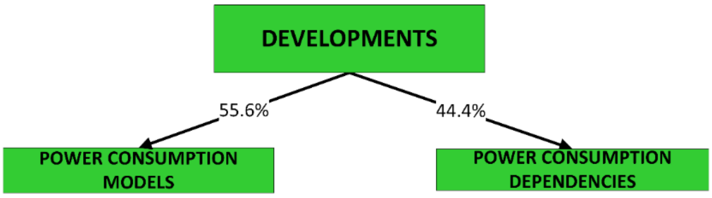


Fig. 43. A top-level division of the developments, with frequency of occurrence shown as percentage

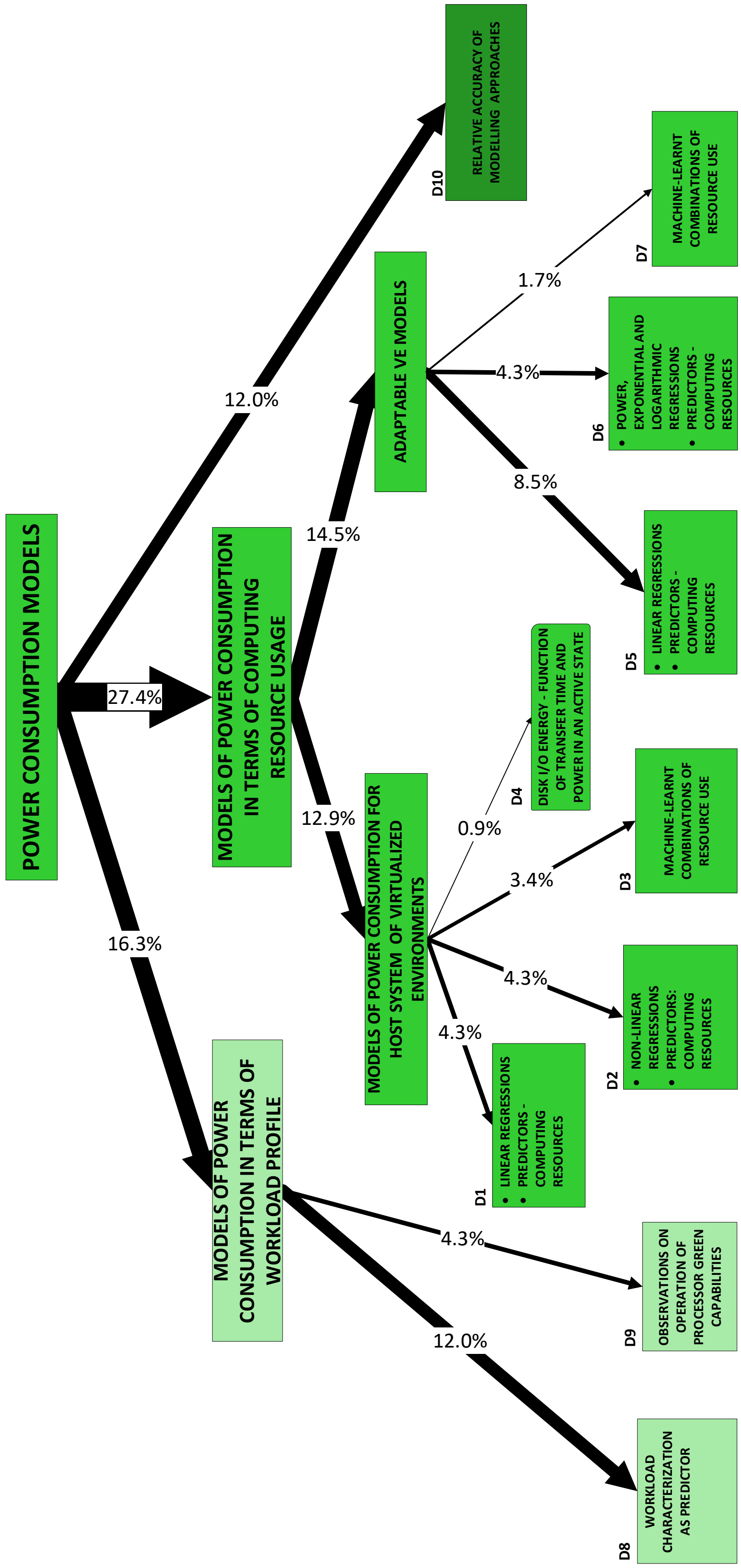


Fig. 44. A taxonomy of power models, with frequency of occurrence shown in line thickness and as percentage

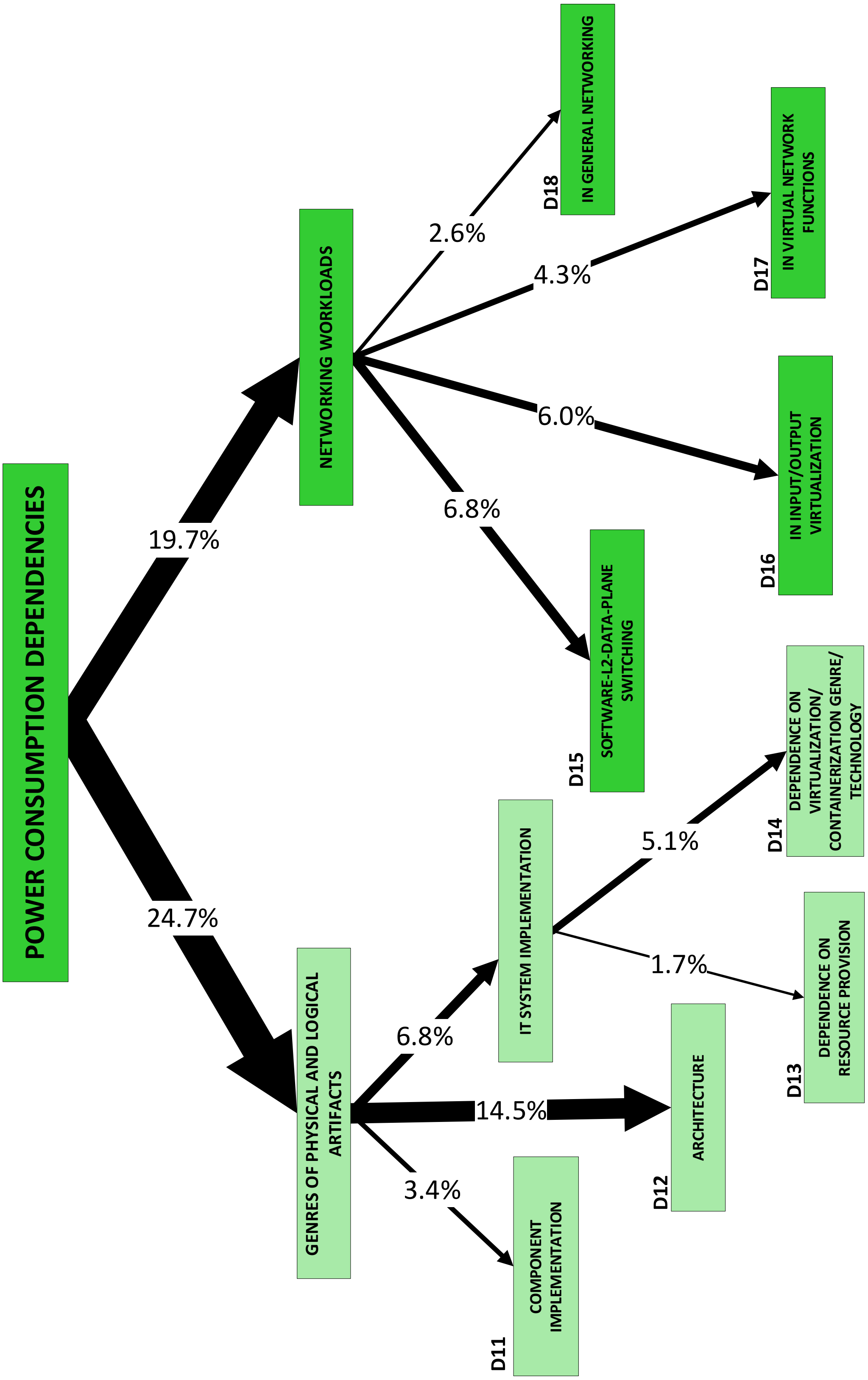


Fig. 45. A taxonomy of power dependencies, with frequency of occurrence shown in line thickness and as percentage

1) Models of power consumption

D1 – D4: I first present four categories of developments that concern ***models of host-system power consumption*** characterized by the condition where ***workload is processed by VCs***:

- D1: linear regression models
- D2: non-linear regression models
- D3: machine-learnt models
- D4: models of local mass storage

This contrasts with the scope of developments referred to under categories D5, D6 and D7, where ***models of VC power consumption*** are presented.

Models in categories D1 – D4 are interesting from the perspective of analyses of ***sets of hosted VCs*** that seek to identify operating conditions of optimal host power efficiency. As predictors, such analyses use instrumentation that measures resources used by the VCs. Categories D1, D2 and D3 all predict power consumption in terms of resource use but differ in the type of model produced.

- D1 regards models of power consumption through linear combinations of scalar predictors [234], [235], [236], [246], [252]. The scalar predictors are resource usage metrics.
- D2 regards polynomial or simple mathematical powers of resource use (the scalar predictors) [234], [237], [238], [253], [258].
- D3 regards models that employ machine learning (e.g., Gaussian Mixture, Support Vector Machine, Neural Networks) [239], [254], [256].

Category D4 regards models of power consumed by mass storage local to the host system [263]. These models attempt to predict power consumption in terms of activity metrics such as total amount of time spent in a known state (in terms of power consumption, e.g., active/idle), rate of data exchange (MB/s or input/output operations per second) and mode of operation (sequential/random and read/write). In the context of the approximations observed in the development of these models, their accuracy cannot be fairly assessed without implementation. My choice of a separate category for this device is not as arbitrary as the first glance might suggest. Few works extend into meaningful consideration of I/O, but as edge computing takes shape, the share of power consumption attributed to I/O devices can fairly be expected to increase.

Adaptable VC models (D5 – D7): Developments within these categories consist of ***adaptable*** models of the virtualization container's power consumption and have two important characteristics in common.

1. They are adaptable to a variable number of concurrent, co-hosted (active on the same host system) VCs (I refer to the latter scope of variability as the seventh dimension of variability).
2. The predictors are the measured amount of computing resources used by the VCs.

Models can be distinguished by the predictors they use, workloads employed and modelling approach:

1. ***predictors*** (of VC power consumption) are obtained from system software's instrumentation, e.g., CPU utilization (see approach categories A10 – A13), and from microarchitecture instrumentation, e.g., LLC misses (again, see approach categories A10 – A13);
2. ***workloads*** used to obtain the model (this restricts the range of workloads within which the model is valid) may be:
    a. specific workloads: the most restrictive, as they relate to a particular test load;
    b. synthetic workloads: less restrictive than specific but limited to exercise of one resource, typically the CPU;
    c. combinational workloads: still less restrictive, involving the exercise of a number of resources of the host system (e.g., SPEC CPU benchmarks may be both processor and memory intensive);
3. representative workloads (e.g. TPC[28]-W [278]) produce models that are readily associated with use cases175odelling***ng approach*** may be:
    a. linear regression (category D5) [235], [236], [246], [250], [252], [258], [259], [261], [263];
    b. power (integer- and non-integer powers), exponential and logarithmic regressions (category D6) [238], [243], [253]
    c. machine-learnt combinations of resource use (category D7) [239], [256]

---

[28] Transaction Processing Performance Council

Models of power consumption that use workload profile as predictors (D8, D9): Categories D8 and D9 group developments from (two) sets of RUs that predict power consumption of hosts and/or VCs through (measurement of) some characteristic of the submitted workload. This contrasts with RUs in categories D1 – D7, where prediction is obtained through (measurement of) some computing resource (processor and/or memory and/or I/O). Most developments in this category are obtained through abstraction of hardware by one or more model parameters that express power consumption under case-specific conditions of operation. Some of these abstractions are identified in the descriptions of these two categories.

Developments in D8 [240], [241], [242], [243], [244], [245], [248], [255], [260], [262], [264], [271] use:

1. processing load (number of processes, millions of instructions per second (MIPS), etc.) pertaining to a specific application, as predictors of ***host system*** power [240], [241], [242], [243], [244], [248]
2. packets per second, through an intrusion detection system implemented in a VNF [262]
3. transcoded frames per second, through a transcoder implemented in a containerized network function (CNF), and inferred images per second, also in a CNF [255]
4. average network transmission rate, as a predictor of host system power [245]
5. statistics of a Batch Markov Arrival Process (BMAP) (packet traffic) as a means of prediction of power consumption by a VNF [260], [271]

Hardware is abstracted through measurement of power consumption at some operating point (a specific operation is being carried out), or change in power consumption over some operating range. Examples follow.

1. In [242], where energy efficiency of an interactive web service is studied, the operating point is an entire VM running the TPC-W benchmark [278].
2. In [244], [248], the operating points are the host's power consumption when (a) idle, (b) one core is active (processing load) and (c) maximum, with all cores active. Furthermore, use is made of the step increment in consumption corresponding to the activation of each additional core. Cores are activated when they are utilized by VCs.
3. In [245], the operating point is the power consumption when co-hosted VCs are transferring a file to a client computer. An ***affine*** relationship between the host's power consumption and its transmission rate (transmissions originate on hosted VCs) is found.

4. In [241], an operating range is used: the increase in power consumption that corresponds to an increase in MIPS on the VCs.
5. In [260], the operating point is the power consumption when a VC running on a single processor core is switching packets at the maximum rate for a given performance state.
6. In [262], fifty-four (54) different features of network traffic are input to an artificial neural network that selects the operating frequency that optimizes power consumption.

With one exception, none of the works in the above list uncovers the hood to peer at the processor's internals (to obtain predictors of power consumption). The exception is [260]; yet even in this case, the performance monitoring counters are not used as direct predictors of power consumption, but to obtain (a) the timing information necessary for a queueing model and (b) the operating state (ACPI (Advanced Configuration and Power Interface [279]) P- and C-states).

Developments in D9 [264] are set within the models branch of the taxonomy. These developments may be considered as useful observations on ***the operation of processors' green capabilities***. Examples of these observations (all from [264]) include:

1. low-power-instructions might be a better candidate than low-power-idle to save power under higher link utilization;
2. operation in full ACPI P-state, operation with low-power-instructions on idle detection and operation with low-power-idle on idle detection are (a) in ascending order of latency to return to active processing and (b) in descending order of power consumption;
3. 80% is a processor utilization threshold below which low packet latency is guaranteed under BMAP traffic arrival.

By "green capabilities" I refer to a broader range of microarchitectural aspects than the by-now-conventional adaptive rate and low-power-idle operation. While these latter two remain at the center of attention, there is also the means of low-power-instructions [264] that has been successfully employed to improve power efficiency. Notwithstanding the origin of these observations in modelling work, it may be argued that they might also be classified within the dependencies branch of the taxonomy. I have chosen the models branch, but as further studies add to the body of knowledge on how to operate processor green capabilities, thi' catego'y's position in the taxonomy might need to be changed.

Relative accuracy of modelling approaches (D10) [237], [239], [240], [254]: Developments presented under category D10 are comparisons of the relative accuracy of alternative modelling methods with respect to conventional polynomial (including linear), power, exponential and logarithmic regression. These developments have been found within works that show models classified under category D9. The purpose is to qualify and quantify improvements of machine-learnt models with respect to conventional regression models.

2) Dependencies of power consumption

This parent node of the taxonomy is divided into two child nodes that are not strictly mutually exclusive. For example, the software data plane is considered in works under D15. Clearly, the software data plane is a logical artifact and might be included within a child node of "physical and logical artifacts" , or directly thereunder as a leaf node (i.e., as a category). The choice of separation of D15 – D18 and inclusion under the parent node "networking workloads" was taken for two principal reasons. Firstly, the recurrence of investigation of power consumption's dependency on networking workload merits attention through separation. Secondly, as this survey caters for an audience with an interest in softwarized networking, an emphasis on power dependency on networking workload seems justified.

Knowledge about dependency of power consumption on specific hardware (D11); dependency of power consumption on architecture (D12); dependency of power consumption on resource provisioning (D13) and dependency of power consumption on virtualization genre & technology (D14): Categories D11 – D14 are grouped into a set of works that obtain the sense of the correlation (positive/negative/none) between power consumption and some genre of artifacts:

1. specific hardware types (D11),
2. computer architecture (D12),
3. resources provided (D13) and
4. virtualization genre and technology (D14).

Category D12 groups works that relate to observations on the impact of architectural features on power consumption [226], [227], [244], [248], [253]. While these observations are useful, they are generally too coarse to be directly applicable to real-time power control. Their use emerges from guidance which they provide in the development of power models. For example, it was observed that when the number of threads that fully occupy a core's time (active 100%) exceeds the number of logical cores in the system, the energy efficiency (measured, in this case, in hash/J, a computational metric of energy efficiency) decreases [227]. Evidently, this is good guidance; equally evidently, it is not a directly applicable development.

Categories D11, D13 and D14 relate closely to (various aspects of) implementation, and as such are of particular interest to the ***system integrator***. Data of high quality from these developments inform and guide the tasks of gathering components into systems that meet the non-functional requirements obtained from concern with energy and power efficiency.

Category D11 gathers observations about the dependency of power consumption on specific processor hardware [226], [246], [253]. I have observed that these developments are gathered as a by-product of the process of research; they are rather incidental. Like any implementation, their usefulness is limited to the lifetime of the concerned device(s).

Category D13 [250] includes developments that regard the specific resource configuration of:

1. the instantiating host, i.e., the relationship between a VC's power consumption and the specific resources of its host hardware specifics such as the number of cores and amount of memory carried by the host instance;
2. the guest VC, i.e., the variation of a VS's power consumption with its resource assignment, on the same host.

Category D14 gathers observations about the dependency of power consumption on instances of virtualization genre and technology [232], [233]. Developments in this category are less incidental than those in category D2 and are obtained with the focused intention of tackling challenges relating power consumption of implementations. These developments relate to a less diversified group of implementations. For example, there are fewer virtualization platforms than processors to choose from. Direct use of these developments is mostly limited to the specific implementations concerned; however, some generalizable conclusions exist. For instance, it was observed that both hardware-assisted virtualization and paravirtualization are less efficient (in the specific empirical setup) than non-virtualized operation in the use of processor caches [233]. This empirical evidence favors the hypothesis that cache hit ratios suffer due to the greater thread rotation in virtualized and containerized environments.

Knowledge about dependency on power consumption while processing a networking workload (D15 – D18): Development categories D15, D16, D17 and D18 are grouped here as their central characterization is knowledge about the behavior of power consumption by VCs while processing a networking workload. Categories D15, D16 and D17 reflect the challenges described in P11, P9 and P10, respectively.

Category D15 regards contributions to knowledge about the power consumption of software data planes [225], [230]; D16 regards virtualization of network I/O [228], [230], and D17 is about network functions [225], [230], [262]. Thus, in [228], the power efficiency of DPDK PMDs is demonstrated with respect to Netmap drivers, for packet transmission. This development is balanced by [225], where the power efficiency of transmission through a DPDK-enhanced Open vSwitch is shown to be worse (@500-byte Maximum Transfer Unit (MTU)) than that of the unenhanced Open vSwitch. In each of these categories, efforts are made to allocate burden through isolation of power consumption and attribution to the sub-system (data plane/virtualized IO/network function) under study.

Category D18's developments differ from those of D15 – D17, because they cut across these latter categories' sub-system boundaries [233], [242]. For example, in [242], the energy efficiency of web transactions executed on co-located VMs is found to be highest in the operating condition of processor-core over-subscription (more VMs than cores). In [233], it is shown that power consumed by packet delivery to a VM through a software packet switch is much higher than that required for delivery in a non-virtualized environment. In both these papers, the object of interest incorporates virtualization of network I/O ***and*** the data plane. In [242], the object of interest encompasses the network function: a web service and accompanying application and database components.

All four categories are of keen interest to the system integrator. Emphasis is on the components in the integrator's set of building blocks, specifically on the behaviour of power consumption of various implementations, and types thereof (in the scope of the categories).

7.3.4 P – Dyads (Problem/Challenge-Approaches) graphics

Over the following pages, I present a comprehensive set of graphics (Fig. 46–Fig. 57) that illustrate the approaches detected to tackle specific challenges. For example, consider Fig. 47. The graphic shows the component approaches applied to discover "the impact of specific architectural attributes of the host system on power consumption of VCs" (see problem category P1).

The presence of approach categories A1, A5, A6, A8 and A10 – A12 does not mean that ***every*** RU tackling this challenge uses a component from ***all*** of the approach categories. It does, however, mean that every RU tackling P1 uses a subset of the approaches shown. Examples (relevant to P1) follow, with references.

1. ***Managed resource provision (A6):*** Prior to running experiments, researchers set the conditions for the experiment through this approach [225], [226], [227], [228].
2. ***Resource-specific workload (A8):*** This may be used to stress the component implementing the architectural attribute under test [225], [226], [227], [228].
3. ***Simple workload characterization and instrumentation (A11):*** The workload must be characterized by some parameter that serves to measure its demand for power [225], [226].

4. ***Resource instrumentation (A10):*** This is an alternative to use of workload profiling (A11) as predictor of power consumption. Rather than use, say, number of threads, or transmit bandwidth (for networking workloads) as predictors, this approach uses resource instrumentation [225], [226].
5. ***Identification of metric of energy efficiency (A5):*** In certain cases [226], [227], energy or power efficiency is investigated, rather than energy or power consumption. In these cases, the researchers identify and use a relevant metric of efficiency, rather than metrics of consumption (watts or joules).

On any of the dyad graphics, the approaches shown (inside the approach categories) include only those which are used by at least one RU that tackles the challenge category that is the root of the dyads. A full tabulation of all approaches in any approach category is delegated to an online repository[29]. Similarly, all individual developments within a category are delegated to the same repository.

---

18. https://github.com/humaira-salam/PowerMeasurementAndModelingRawData

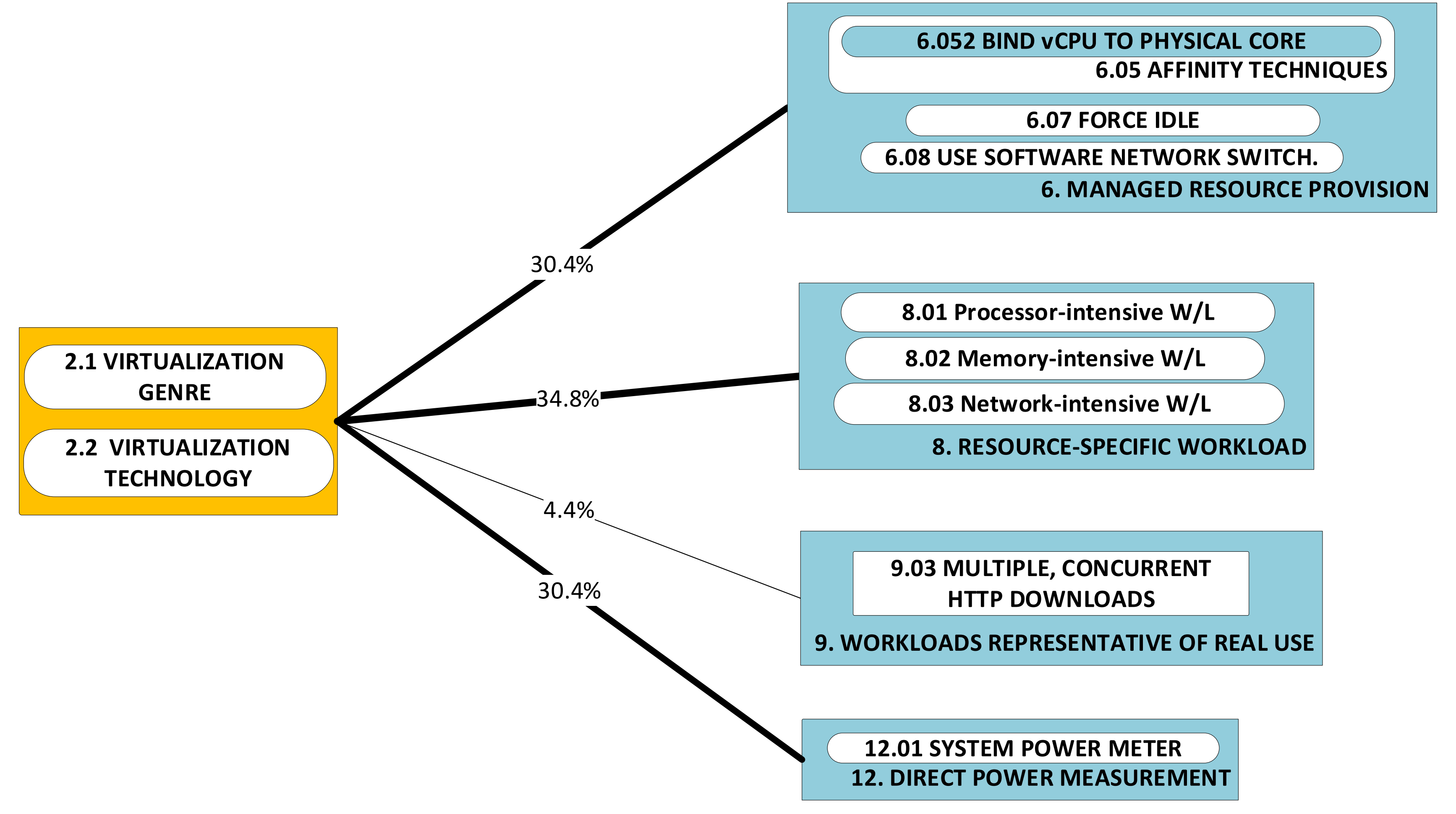


Fig. 46. Approaches to solving challenges in category P2; utility metric U_{A_k} is shown in line thickness and as percentage.

RUs in the respective categories are the following: P2 [232], [233], **A6** [232], [233], **A8** [232], [233], **A9** [232], [233], **A12** [232], [233]

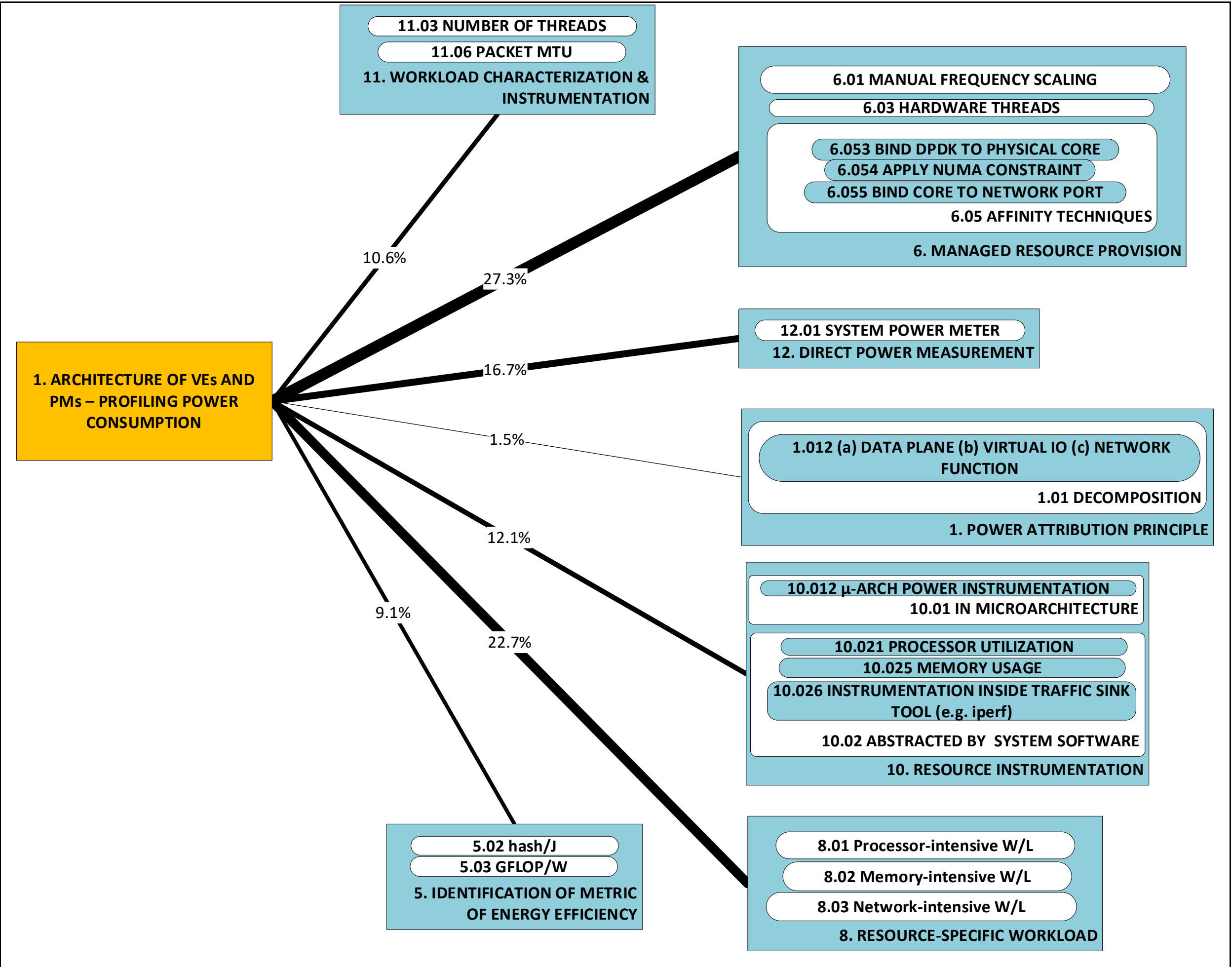


Fig. 47. Approaches to solving challenges in category P1; utility metric U_{A_k} is shown in line thickness and as percentage;

RUs in the respective categories are the following: P1 [225], [226], [227], [228], **A1**[225], **A5**[226], [227], **A6**[225], [226], [227], [228], **A8**[225], [226], [227], [228], **A10**[225], [226], **A11**[225], [226], **A12**[225], [226], [227].

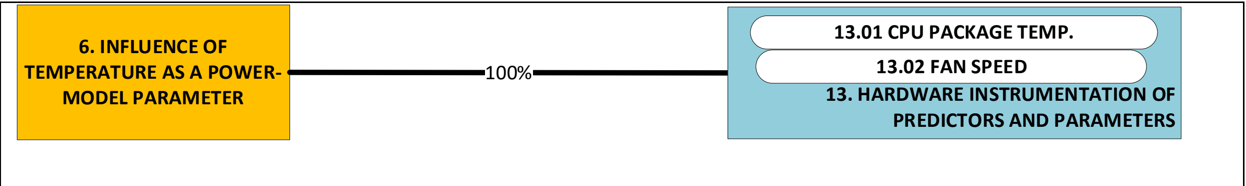


Fig. 48. Approaches to solving challenges in category P6; utility metric U_{A_k} is shown in line thickness and as percentage.

RUs in the respective categories are the following: P6 [237], [258], [259]; **A13** [237], [258], [259].

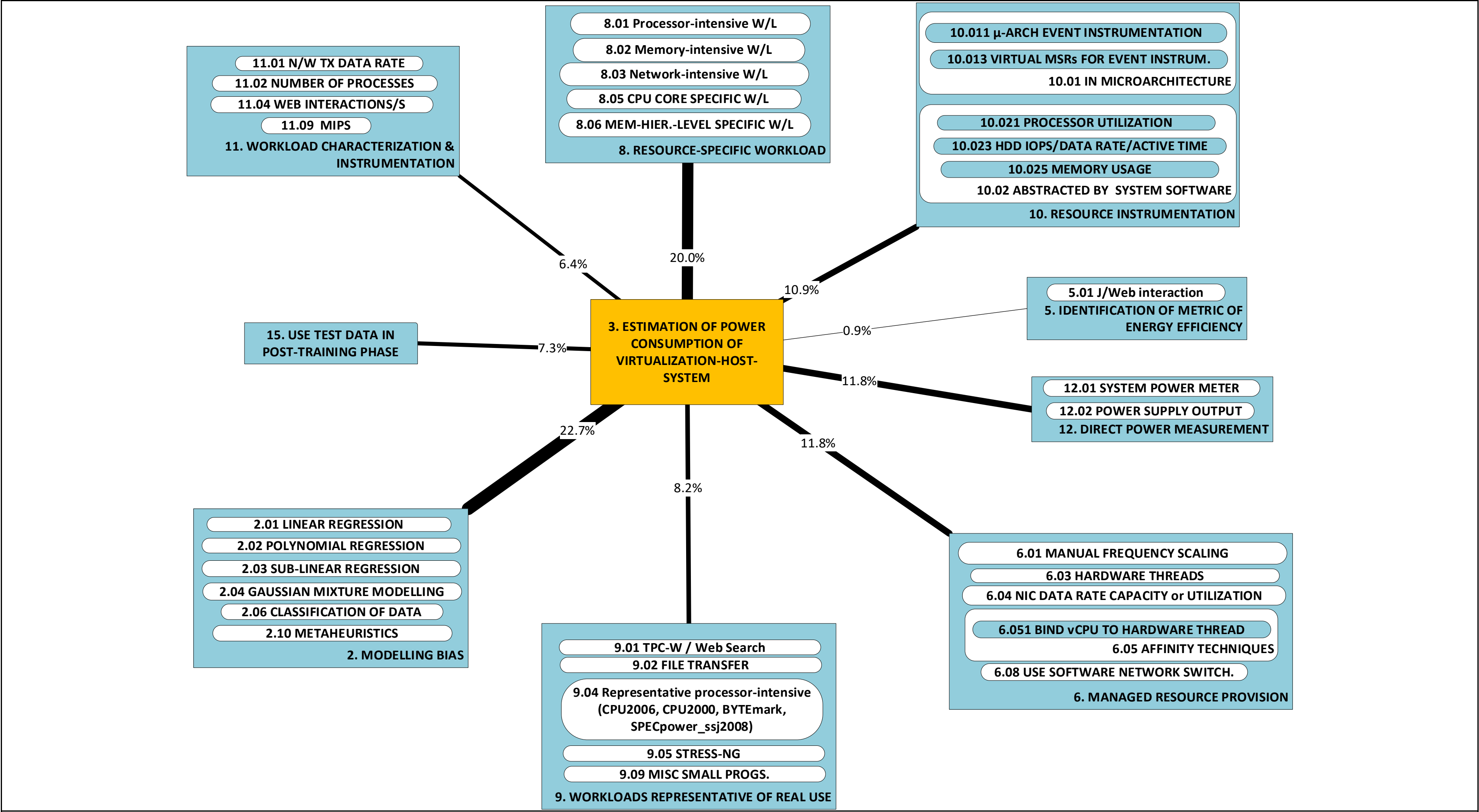


Fig. 49. Approaches to solving challenges in category P3; utility metric U_{A_k} is shown in line thickness and as percentageApproaches to solving challenges in category P3; utility metric is shown in line thickness and as percentage

RUs in the respective categories are the following: P3 [234], [235], [236], [237], [238], [239], [240], [241], [242], [243], [244], [245], [246], [248]**; A2** [234], [235], [236], [237], [238], [239], [240], [241]**; A5** [242]**; A6** [237], [243], [244], [245]**; A8** [236], [237], [238], [240], [241], [243], [244], [245], [248]**; A9** [234], [235], [237], [239], [242], [245]**; A10** [234], [235], [236], [238], [239], [240], [241], [246]**; A11** [241], [242], [244], [245], [248]**; A12** [234], [235], [236], [238], [239], [240], [241], [244], [245], [248]**; A15** [234], [239], [240], [241]**.**

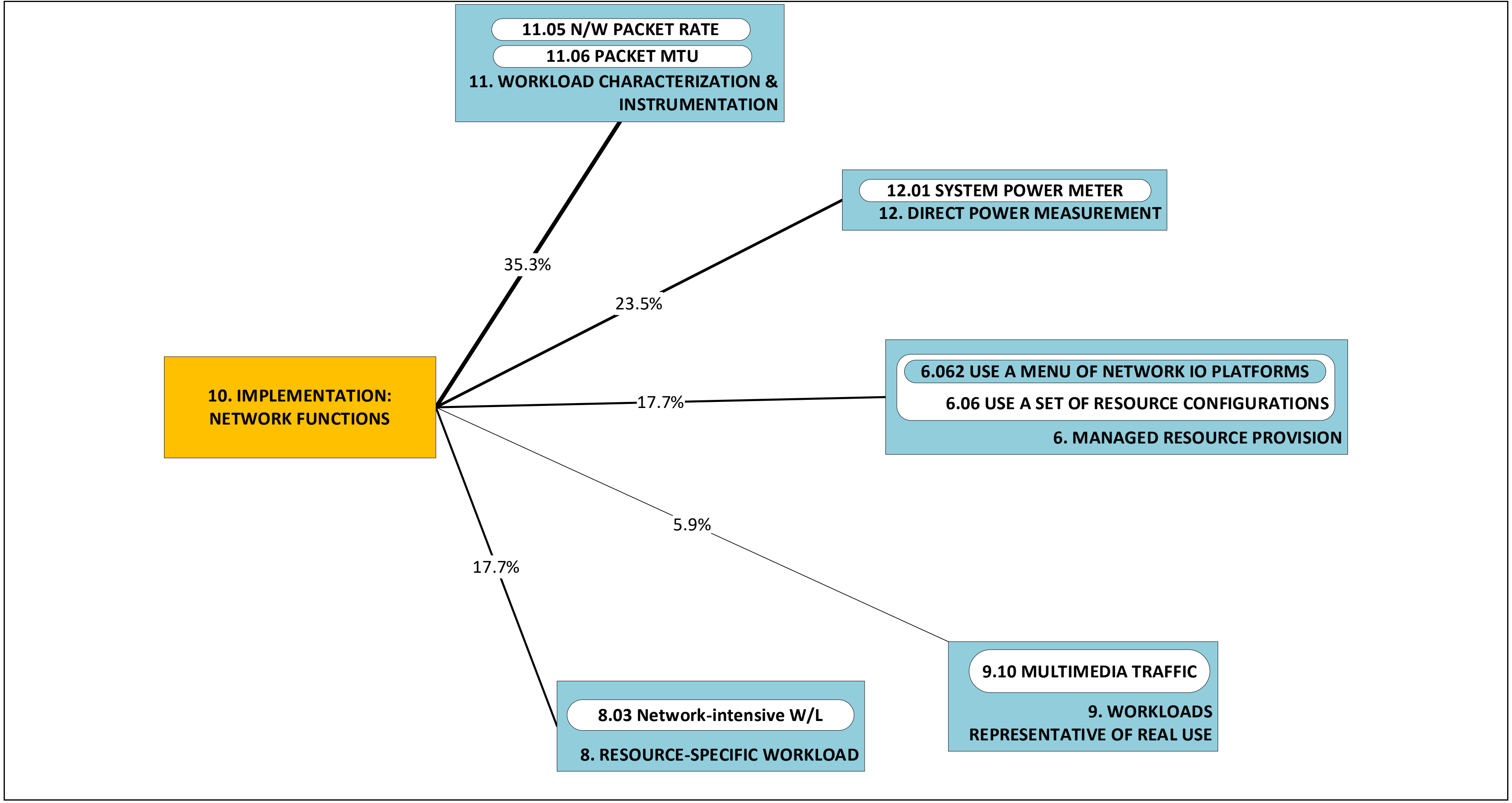


Fig. 50. Approaches to solving challenges in category P10; utility metric U_{A_k} is shown in line thickness and as percentage

RUs in the respective categories are the following: P10 [225], [230]; **A6** [230]; **A8** [230]; **A9 [48]; A11** [230]; **A12** [225], [230].

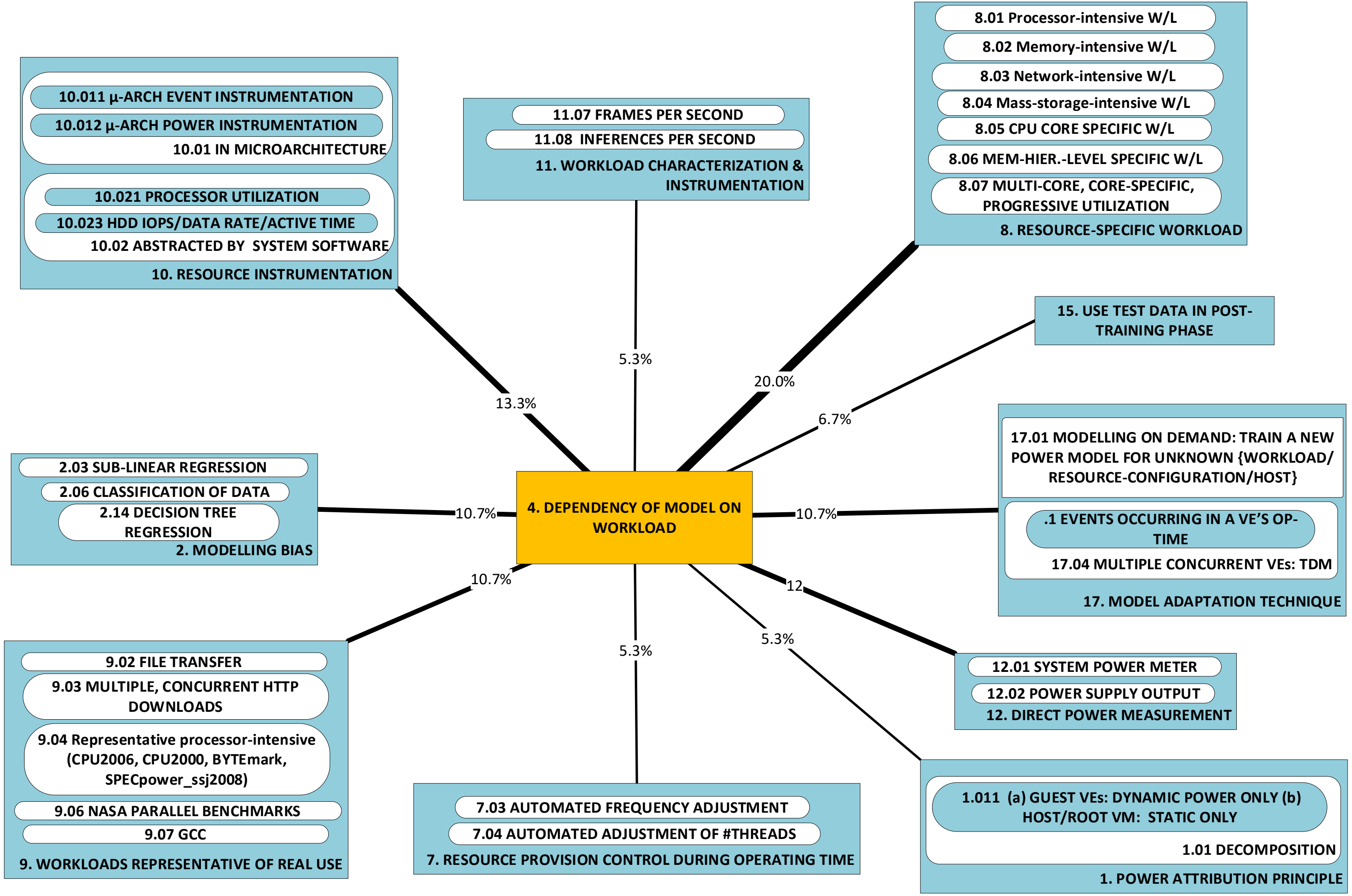


Fig. 51. Approaches to solving challenges in category P4; utility metric U_{A_k} is shown in line thickness and as percentage. RUs in the respective categories are the following:

P4 [233], [239], [244], [245], [246], [248], [252], [253], [254], [255], [256]; **A1** [239], [254]; **A2** [239], [254]; **A7** [255]; **A8** [233], [244], [246], [248], [252], [253]; **A9** [233], [239], [245], [252], [256]; **A10** [239], [255], [256]; **A11** [255]; **A12** [233], [239], [253], [256]; **A15** [239], [254]; **A17** [239], [244], [245], [248], [252], [253], [255].

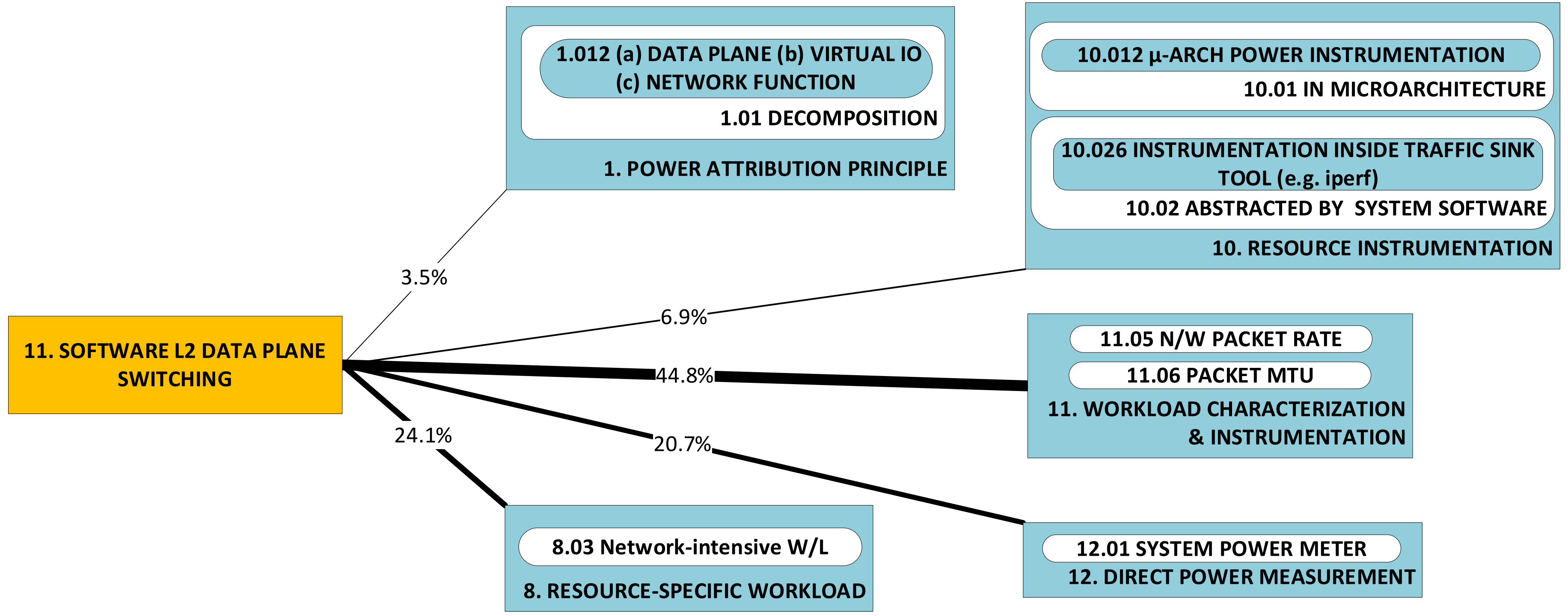


Fig. 52. Approaches to solving challenges in category P11; utility metric U_{A_k} is shown in line thickness and as percentage.

P11 [225], [230]; **A1** [225]; **A8** [225], [230]; **A10** [225]; **A11** [225], [230]; **A12** [225], [230].

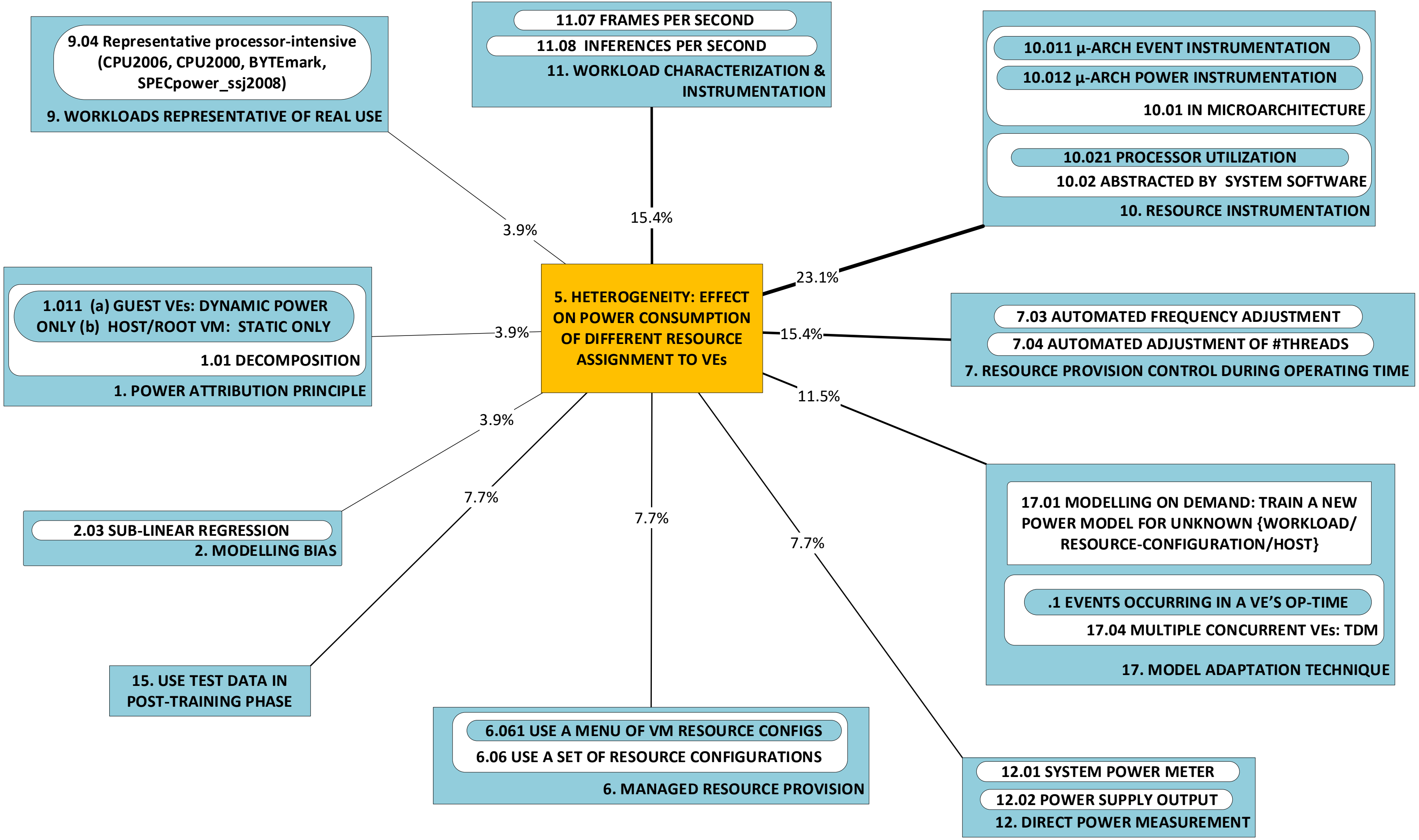


Fig. 53. Approaches to solving challenges in category P5; utility metric U_{A_k} is shown in line thickness and as percentage.

P5 [239], [249], [255]; **A1** [239]; **A2** [239]; **A6** [249]; **A7** [255]; **A9** [239]; **A10** [239], [255]; **A11** [255]; **A12** [239], [249]; **A15** [239]; **A17** [239], [255].

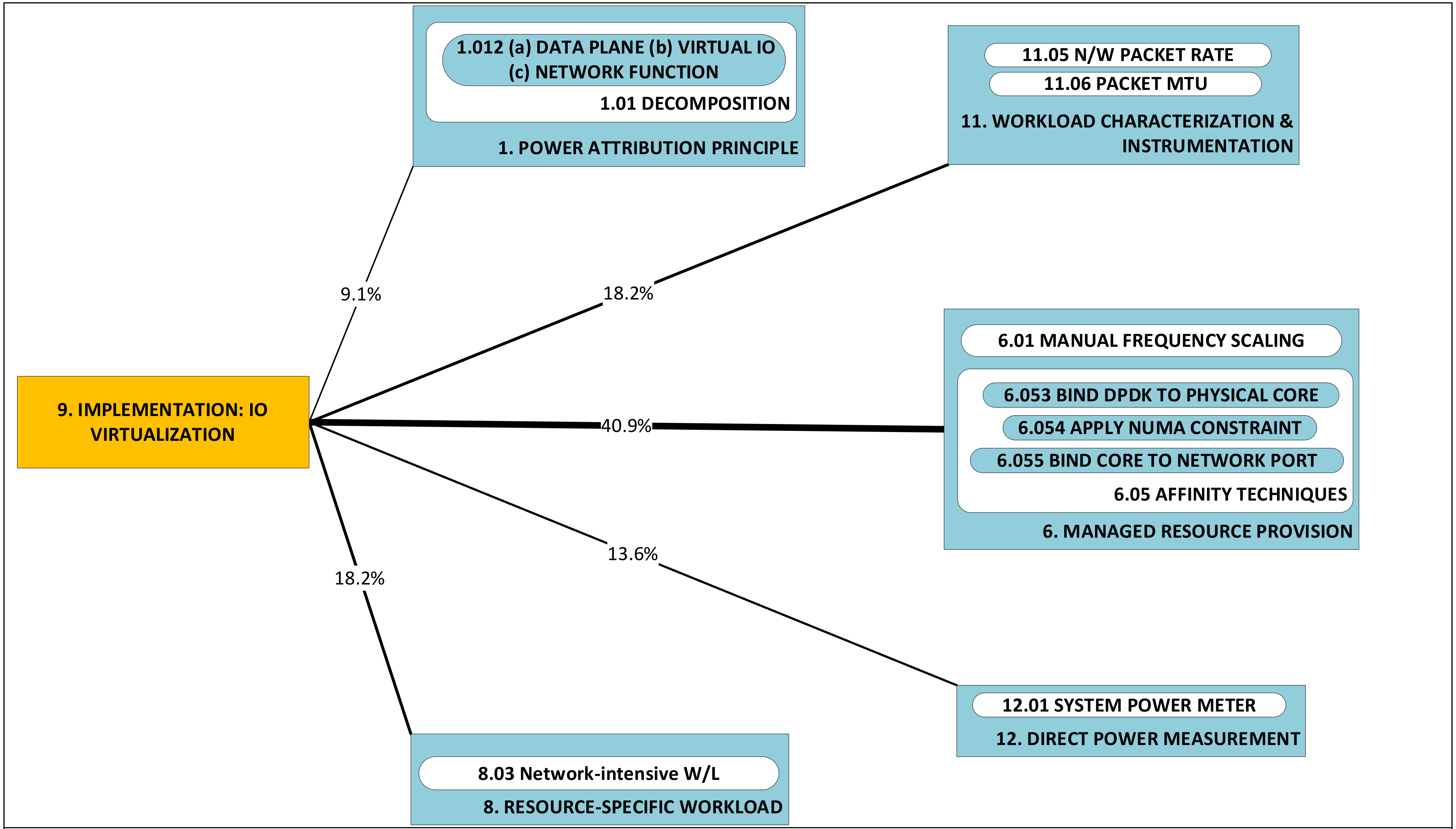


Fig. 54. Approaches to solving challenges in category P9; utility metric U_{A_k} is shown in line thickness and as percentage.

P9 [228], [230]; **A1** [230]; **A6** [228]; **A8** [228], [230]; **A11** [230]; **A12** [230].

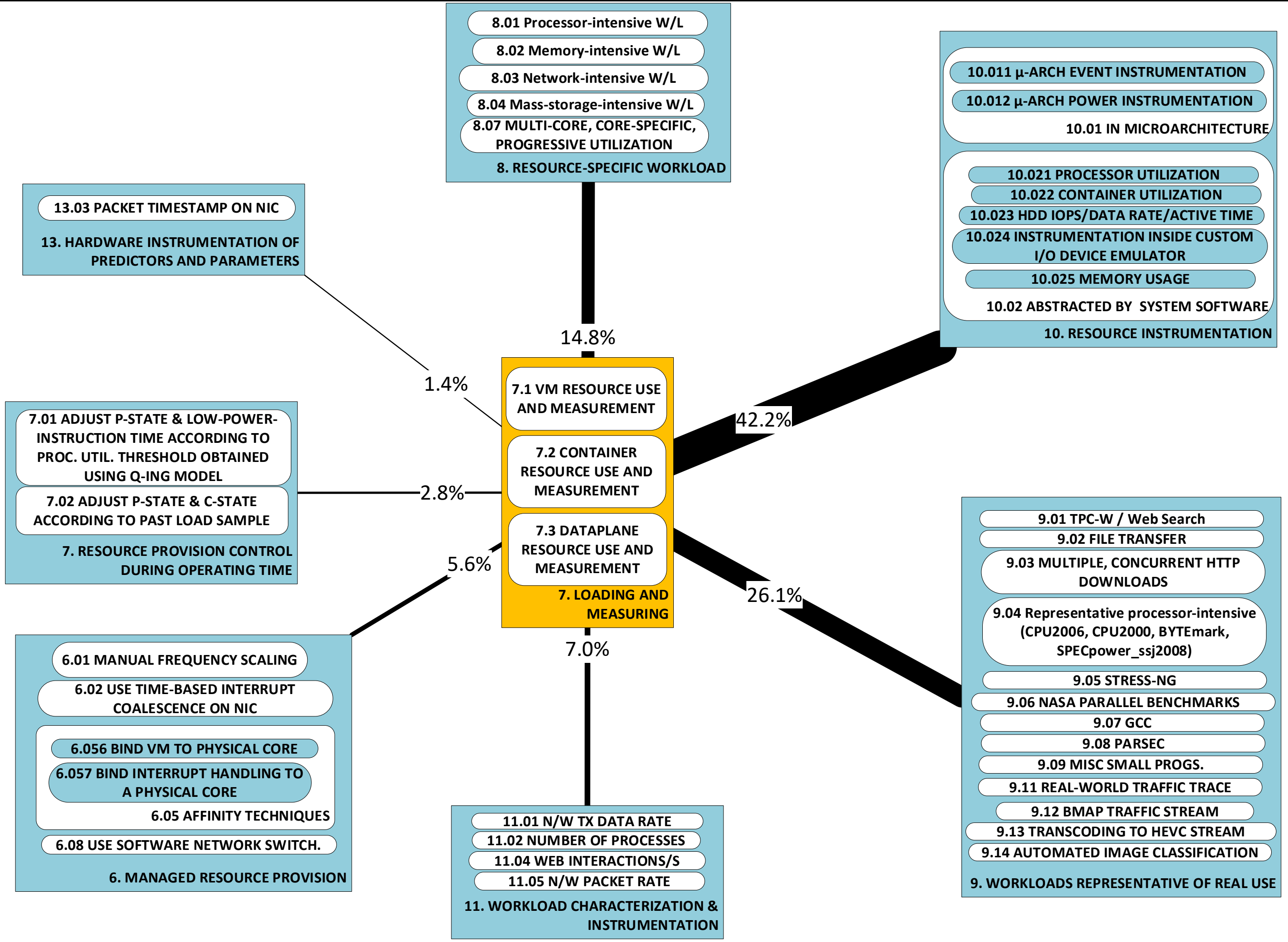


Fig. 55. Approaches to solving challenges in category P7; utility metric U_{A_k} is shown in line thickness and as prcentage

P7 [233], [234], [235], [236], [238], [239], [242], [245], [246], [248], [249], [252], [253], [254], [255], [256], [258], [259], [260], [261], [262], [264]**; A6** [233], [245], [258], [259], [260]**; A8** [233], [236], [238], [248], [253], [254], [261], [262]**; A9** [233], [234], [235], [239], [242], [245], [246], [249], [255], [256], [258], [259], [260], [261], [264]**; A10** [233], [234], [235], [236], [238], [239], [246], [249], [252], [253], [254], [255], [256], [258], [259], [260], [261], [262], [264]**; A11** [242], [245], [248], [260], [261], [264]**; A13** [264]**.**

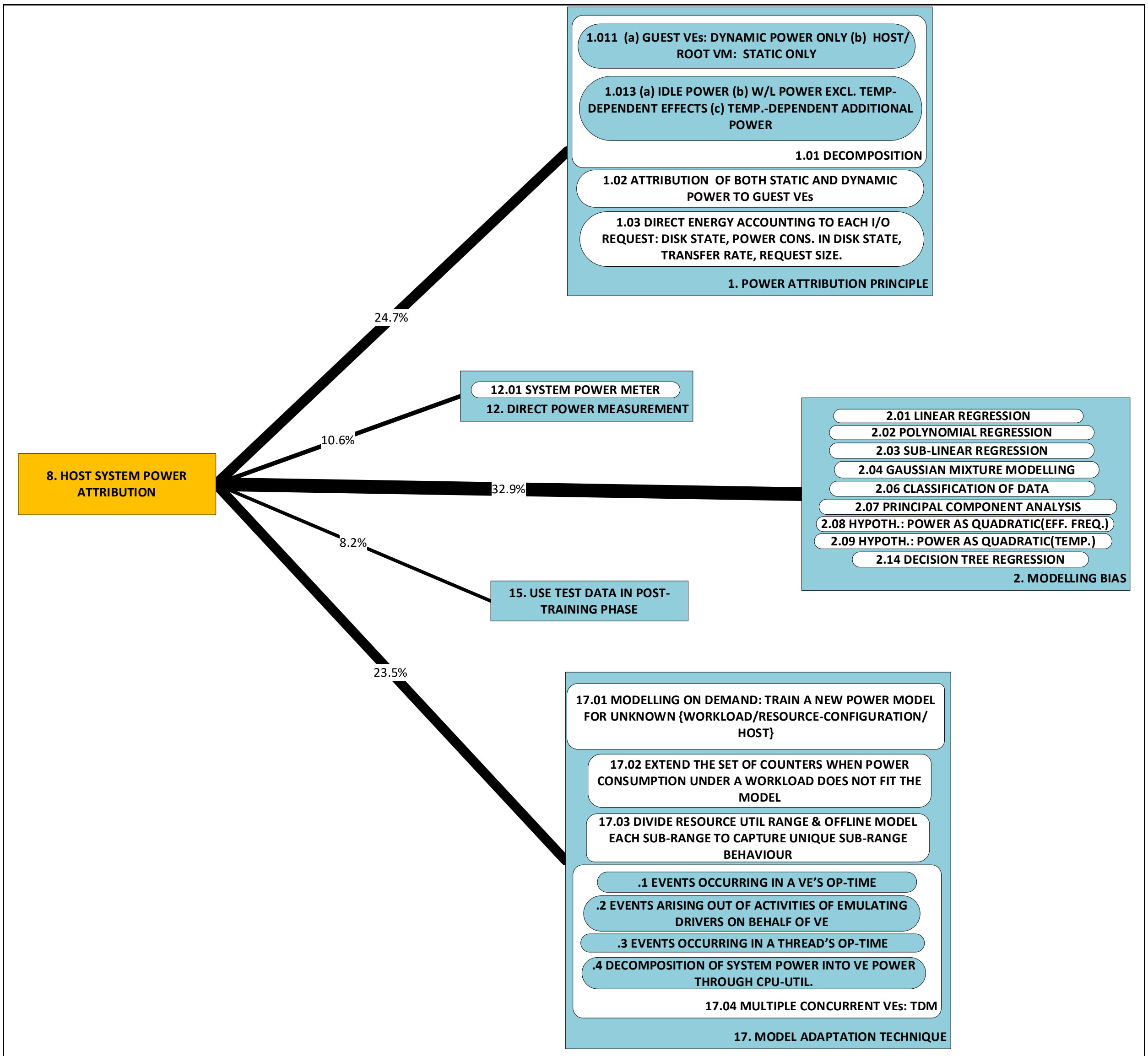


Fig. 56. Approaches to solving challenges in category P8; utility metric U_{A_k} is shown in line thickness and as percentage

P8 [234], [235], [236], [238], [239], [246], [249], [252], [253], [254], [256], [258], [259], [261], [263];

A1 [234], [235], [236], [238], [239], [246], [249], [252], [253], [254], [256], [258], [259], [261], [263];

A2 [234], [235], [236], [238], [239], [249], [252], [253], [254], [256], [258], [261];

A12 [234], [235], [236], [238], [249], [253], [261];

A15 [234], [239], [254];

A17 [234], [235], [236], [238], [239], [246], [249], [252], [253], [256], [259], [261], [263].

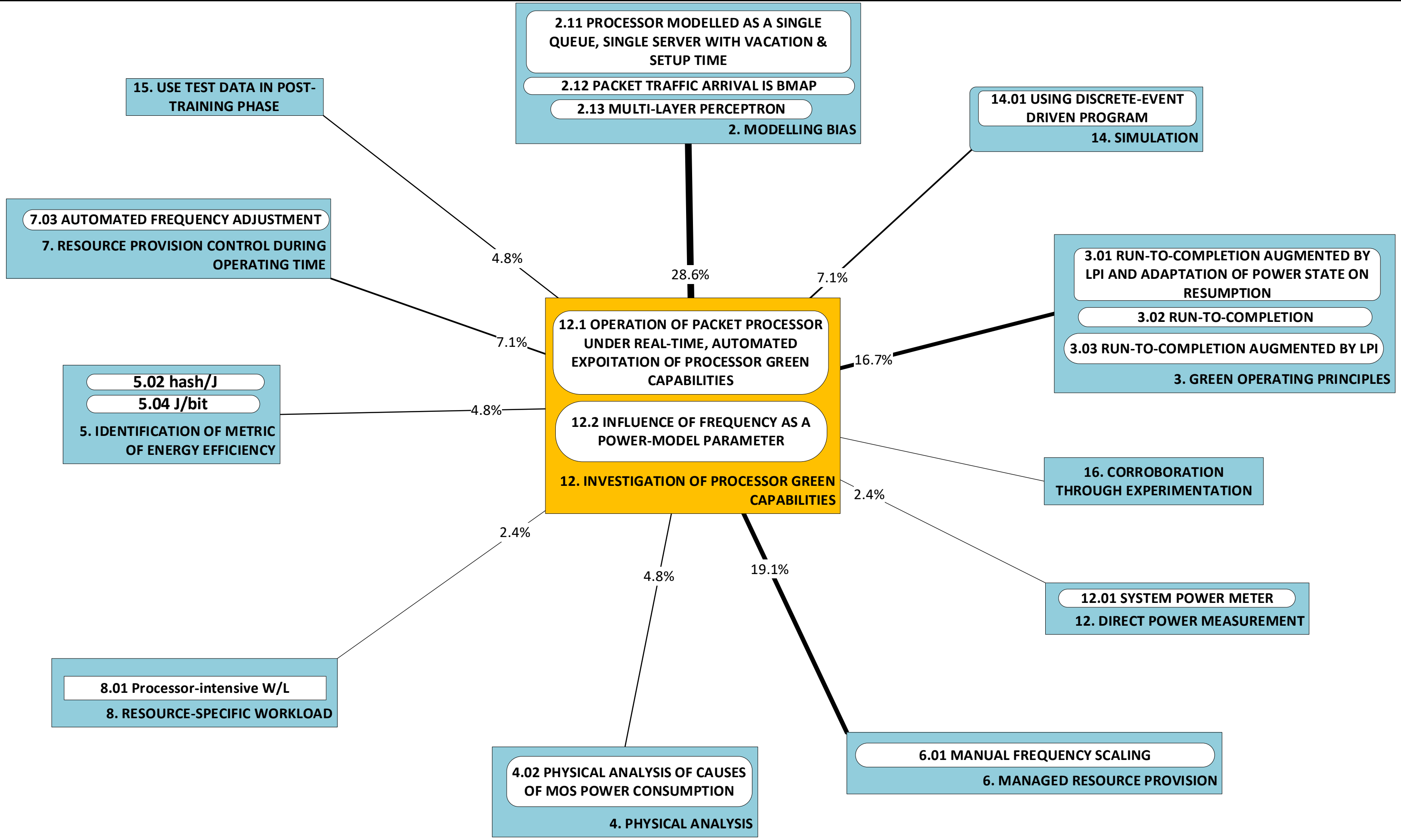


Fig. 57. Approaches to solving challenges in category P8; utility metric U_{A_k} is shown in line thickness and as percentage

P12: [227], [237], [244], [258], [259], [260], [262], [264], [271]; **A2** [260], [262], [264], [271]; **A3**[260], [264], [271]; **A4** [237]; **A5** [227], [262]; **A6** [227], [237], [244], [258], [259]; **A7** [262]; **A8** [227] ; **A12** [227] ; **A14** [264]; **A15** [262]; **A16** [260].

7.3.5 Causality DAG

The Causality Directed Acyclic Graph (DAG) in Fig. 58 shows a bird's eye-view of the proceedings of research in scope.

7.3.6 Triads (Problem/Challenge-Approaches-Developments) Graphic

Fig. 59 shows the triads graphic. To improve readability, illustration is limited to the triads that are in the top twenty percentile of a list ordered according to thickness. These triads comprise 49.2% of the total number of compiled triads.

7.3.7 Statistics

Bar charts that illustrate the category metrics described in [223] (see section on “Statistics”) are presented below.

1. Challenges (Fig. 60)
    a. frequency of occurrence in the set of all RUs;
    b. frequency of occurrence in the set of all challenges in all RUs;
    c. frequency of occurrence, weighted by approach diversity, in the set of all challenges in all RUs.
2. Approaches: (Fig. 61)
    a. frequency of occurrence in the set of all approaches in all RUs;
    b. frequency of occurrence in the set of all triads in all RUs.
3. Developments: frequency of occurrence in the set of all developments in all RUs: (Fig. 62)

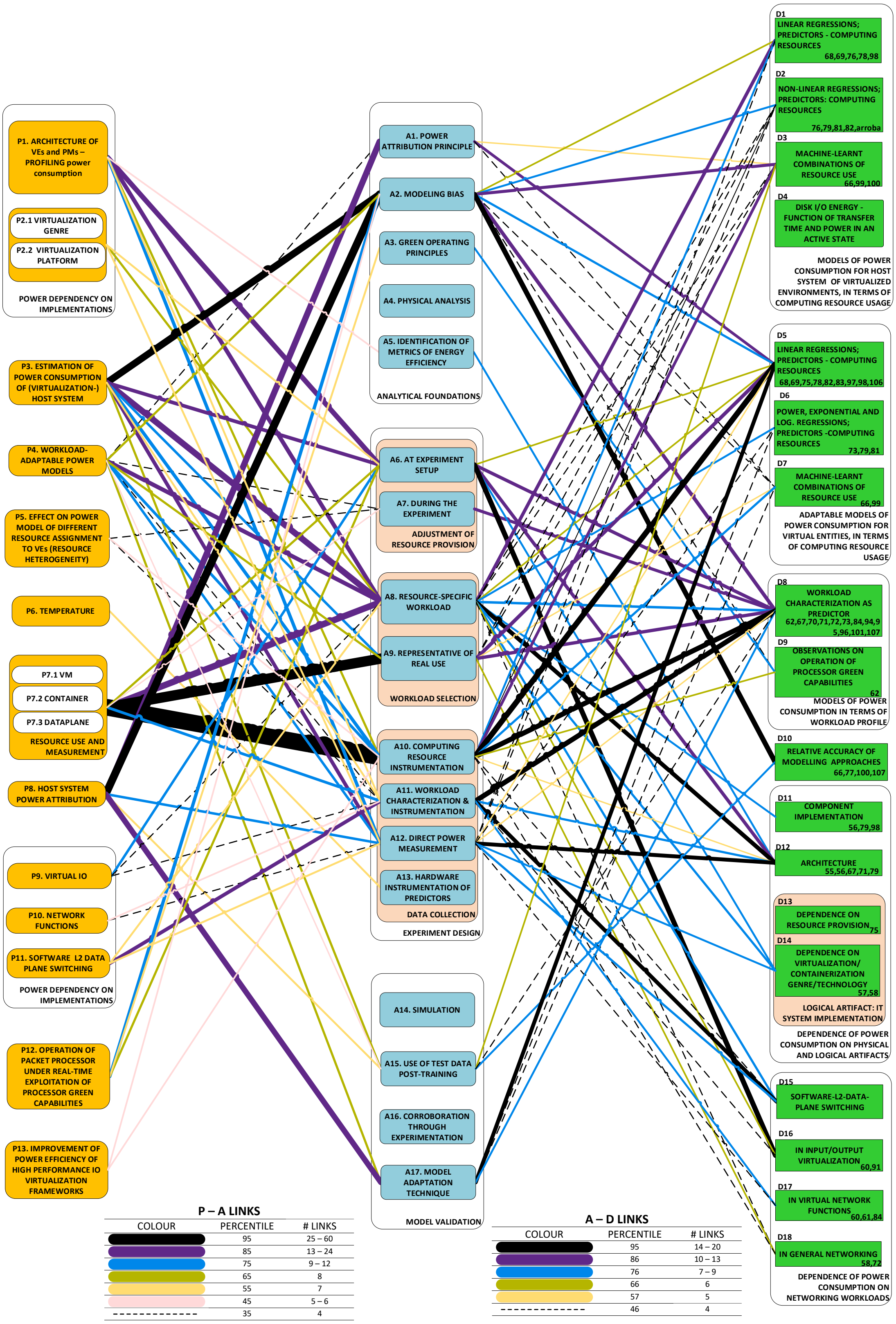
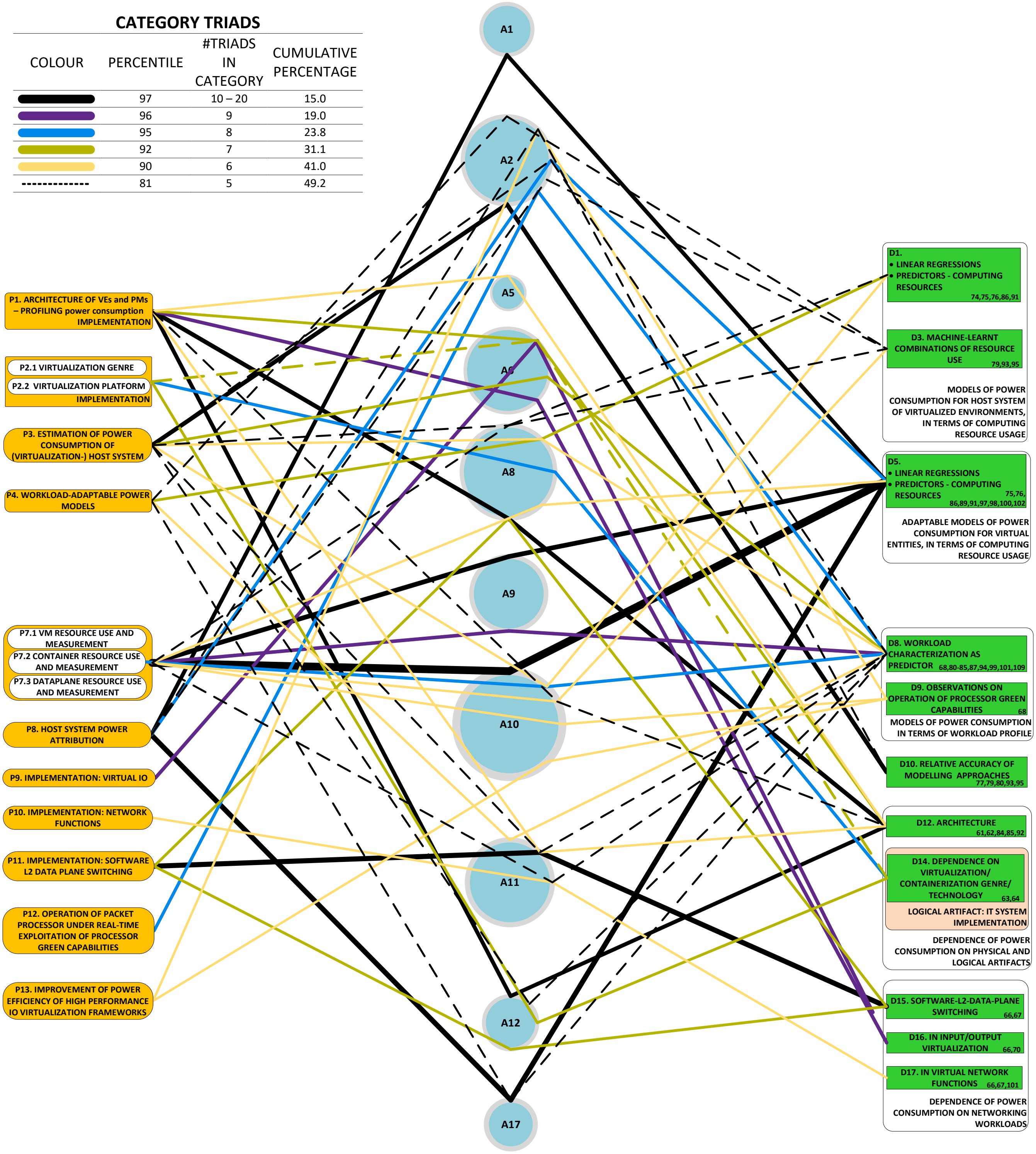


Fig. 58. A directed acyclic graph showing the distribution of research into power measurement & power consumption models in virtualized networking and computing environments



COLOUR	PERCENTILE	#TRIADS IN CATEGORY	CUMULATIVE PERCENTAGE
▬	97	10 – 20	15.0
▬	96	9	19.0
▬	95	8	23.8
▬	92	7	31.1
▬	90	6	41.0
-------------	81	5	49.2

Fig. 59. Triads

To improve legibility, the x-axis labels show category numbers only. The codes ("terse, dense representations of a verbose articulation of a concept", see [223, N. see sub-section 'what are codes?']) linked to the numbers are shown in Table XII , Table XIII and Table XIV . Table XII also shows questions that help to clarify articulation of the challenge posed.

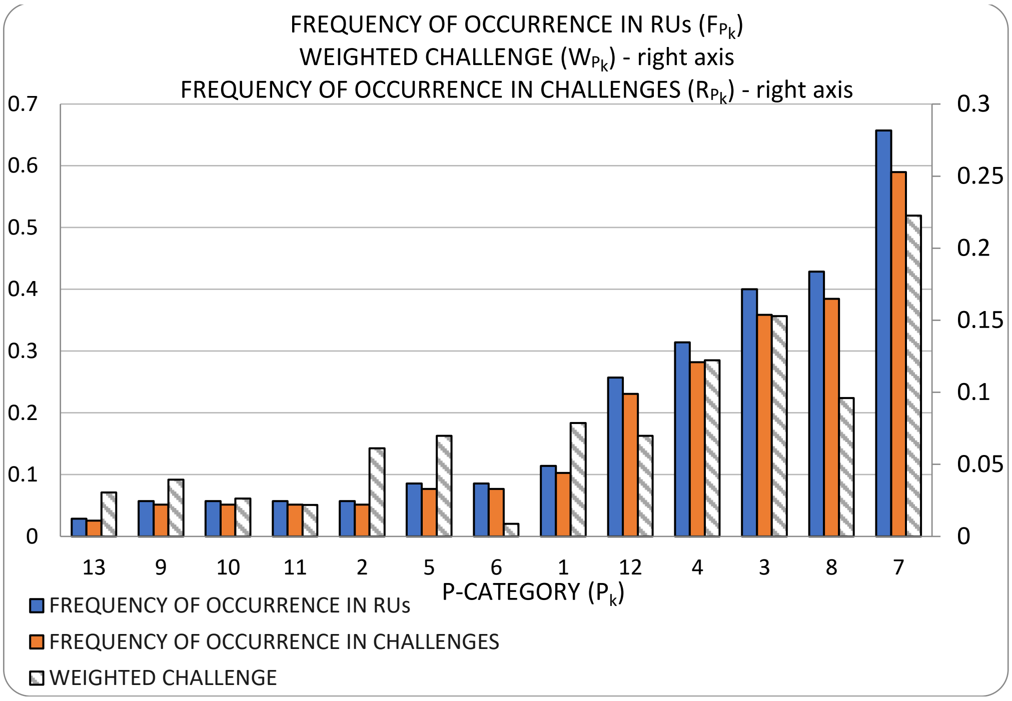


Fig. 60. **Frequency, Research Interest and weighted Challenge bar chart**

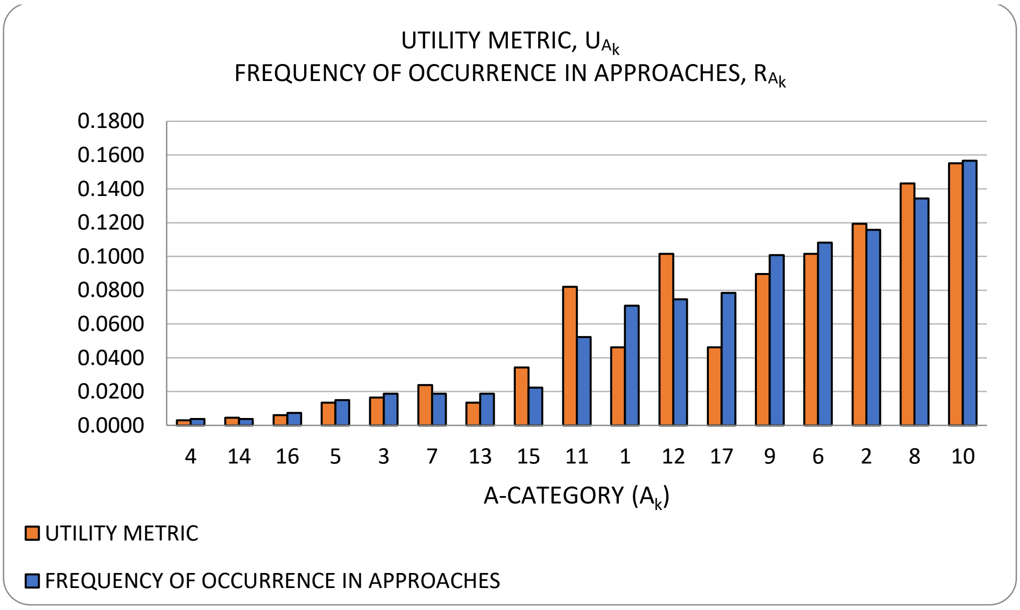


Fig. 61. **Approach metrics**

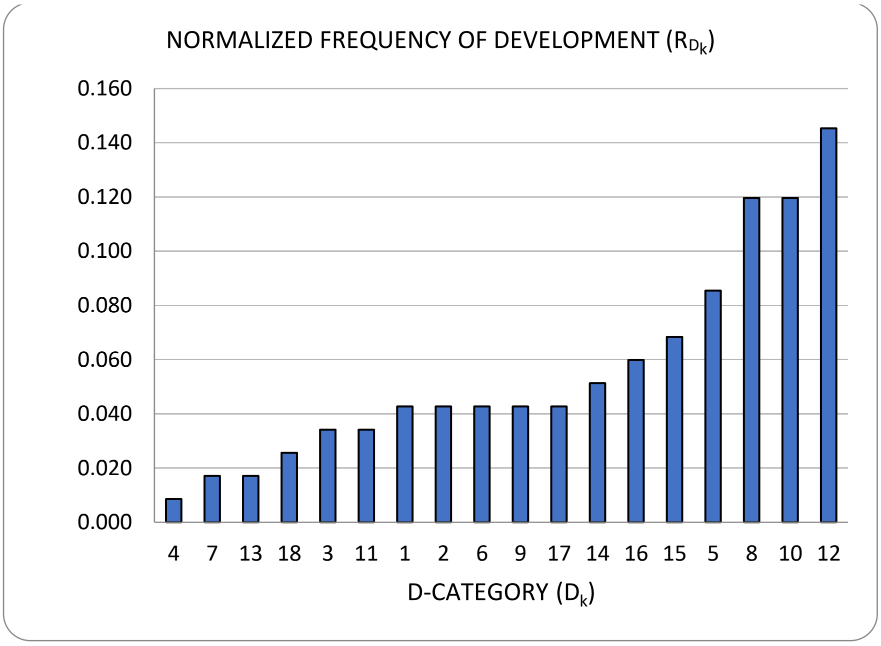


Fig. 62. Normalized frequency of Developments

Table XII CHALLENGE CATEGORY CODES AND REPRESENTATIVE CHALLENGE QUESTIONS.

CAT #	CODE	PROBLEM
P1	Architecture of VCs and PM– - profiling power consumption	What is the impact of specific major architectural features of computer hardware on the power consumption of VCs?
P2	Virtualization genre, platform	Is it possible to meaningfully rationalize the behavior of VC power consumption across different implementations of systems for virtualization?
P3	Estimation of power consumption of virtualization-host systems	Can host system power consumption be predicted on the basis of VC activity?
P4	Workload-adaptable power models	Can workload-adaptable power models be developed?
P5	Resource heterogeneity	How do a VC's power consumption and power model vary with resource configuration (heterogeneous VCs) ?
P6	Influence of temperature on power-model	How does temperature of operation affect VCs' power models?
P7	Resource use and measurement of VCs	How can load be targeted at specific VC resources? How can the actual resource consumption be measured?
P8	Host system power attribution	How can the (measured) power consumption of a host be attributed to the hosted VCs?
P9	IO Virtualization	What is the impact of specific major implementations of IO virtualization on the power consumption of VCs?
P10	Network functions	Which particular implementation of a network function is most power or energy efficient?

P11	Software L2 data plane switching	What is the impact of specific major implementations of software layer 2 data plane switching on the power consumption of VCs?
P12	Investigation of processor green capabilities	How can we model operation under real-time exploitation of processor green capabilities?
P13	Improvement of power efficiency of high-performance IO virtualization	How can we improve the power efficiency of high-performance packet IO frameworks?

Table XIII APPROACH CATEGORY CODES

CAT #	CODE
A1	Power attribution principle
A2	Modeling bias
A3	Green operating principles
A4	Physical analysis
A5	Identification and use of metrics of energy efficiency
A6	Managed resource provision (setup time)
A7	Controlled resource provision (operation time)
A8	Resource-specific workloads
A9	Workloads representative of real use
A10	Computing resource instrumentation
A11	Workload characterization and instrumentation
A12	Direct power measurement
A13	Hardware instrumentation of predictors
A14	Simulation
A15	Use test data in post-training phase
A16	Corroboration through experimentation
A17	Model adaptation technique

Table XIV DEVELOPMENT CATEGORY CODES

CAT #	CODE
D1	Host models -> linear regressions: predictors = computing resources
D2	Host models -> non-linear regressions: predictors = computing resources
D3	Host models -> machine-learnt: inputs = computing resources
D4	Host models -> mass storage energy consumption
D5	Adaptable VC models -> linear regressions: predictors = computing resources
D6	Adaptable VC models -> power, exponential and log regressions: predictors = computing resources
D7	Adaptable VC models -> machine-learnt: inputs = computing resources
D8	Host/VC models of power consumption -> predictors = workload characteristics
D9	Host/VC models of power consumption- > observations on operation of processor green capabilities
’D18	Pow’r's dependencies -> networking workloads -> in general networking

7.4 Thematic analysis

The themes presented in this section are the product of a thematic analysis undertaken according to the method described in [223]. I first present an overview through a graphic (Fig. 63) that groups the themes and then proceed to an exposition of the themes within the sub-sections.

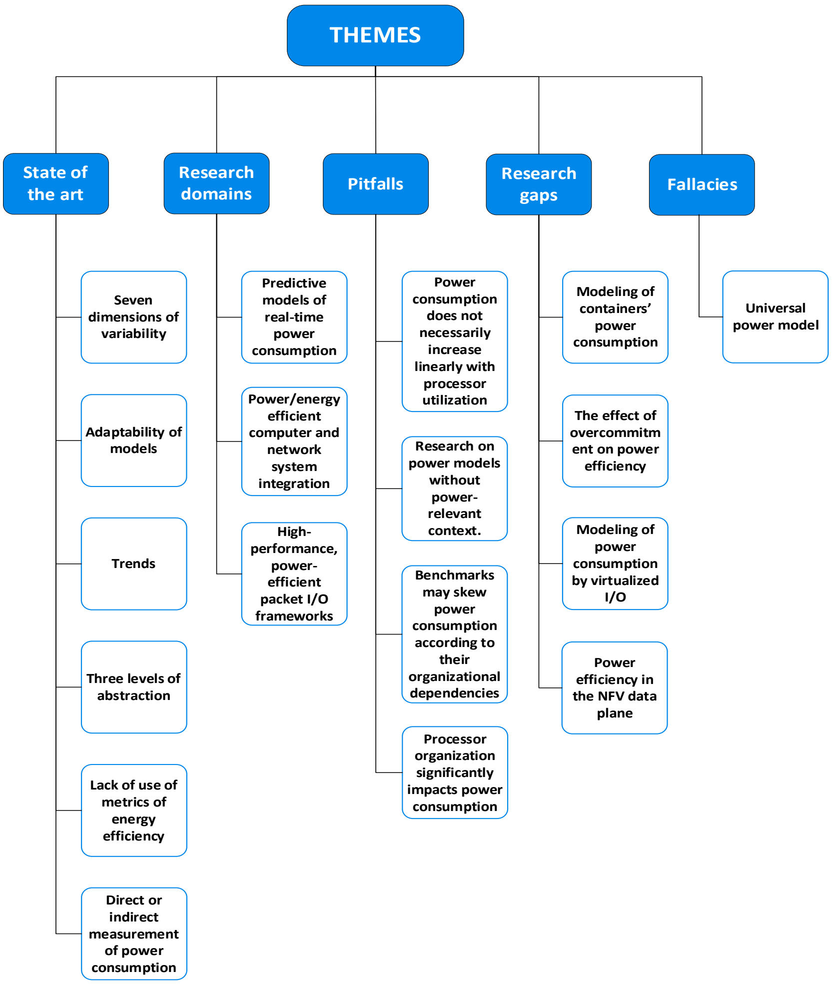


Fig. 63. **A graphical overview of the thematic analysis of the research space**

7.4.1 State of the art

1) Seven dimensions of variability

Through collation and resolution of the derivatives of the core challenge in this research space, seven dimensions of variability in modelling power consumption by a virtual entity were observed. Comparison with the power consumption of physical machines throws the core challenge into sharper relief. With physical machines:

- power consumption can be measured directly;
- there is no virtualizing agent to consider, and
- the activity of other physical machines (that do not send or receive workload) is irrelevant.

The seven dimensions of variability are shown in Table XV . It is not surprising that the problems I have seen researchers tackle are closely aligned with these dimensions. The research space is precisely about the need to obtain an understanding of the impact that these variables have on power consumption. The scope of models found in the RUs covers only a subspace of the seven-dimensional space but the extent is not usually stated. This issue is touched upon briefly in treatment of the pitfall of research without context and the fallacy of the universal power model.

Table XV SEVEN DIMENSIONS OF VARIABILITY

DIMENSION	PROBLEM CATEGORY
1. Workload type	P4, P7
2. Virtualization agent	P2, P9, P11, P13
3. Host (resources and architecture)	P1, P3
4. Temperature	P6
5. Power attribution	P8
6. Co-hosted, concurrent VCs	P8, P5
7. Frequency	P12, P13

2) Adaptable models

Three conditions were observed, which must be met for an automated modelling system to obtain a model of VC power consumption.

1. The first, fundamental condition is common to all successful research into modelling of VC power consumption. ***Resources utilized by the VC*** (whether measured through architectural or microarchitectural instrumentation) ***must be strongly correlated to power consumption by the VC as well as power consumption in host system overheads***. If the overheads are uncorrelated to the VC's activity, or weakly so, then any significant power overhead must be modelled through separate consideration of its driving causes, e.g., temperature [258].

Furthermore, in order that modelling may apply across a diversity of operating contexts, two other conditions must hold.

2. Any change in the parameters of correlation must be dynamically learnt and the model adjusted.
3. Any change in operating context that invalidates parameters of correlation, must be of finite duration. An indefinite transient precludes the formation of a model.

To this observation on the three conditions, I add another observation. Two of the seven dimensions of variability are commonly investigated in validation of the accuracy of modelling systems: ***workload type and co-hosted, concurrent VCs.*** I suggest that a modelling system may be labelled adaptable if, minimally***, it meets the above three conditions under these two dimensions, i.e. (a) operations of variable workload type per VC, and (b) a variable number of concurrent, co-hosted VCs.*** Before proceeding to refer to validations observed in the RUs, it is useful to draw attention here to the need for a ***modelling system***, rather than simply a model, in the estimation of power consumption. Dynamic adjustments are effected through the intervention of such a system, which adjusts model parameters to the operating context. I now proceed to describe how the variables in the two dimensions were varied in some studies.

Workload type: Different types of workload correspond to different use of resources; hence, behaviour of power consumption also differs. Several researchers [225], [252], [253] studied the effect of changing workload type either by categorizing the workload itself or by categorizing the resources consumed by a specific kind of workload. Both [252] and [253] characterized the use of CPU utilization for workloads, which is known to have a workload-type dependent relationship with power consumption [239]. Thus, in both these studies, the need arises to re-train on change of workload type (second condition). Further, both carry an observation about the duration of re-training for refined models (third condition).

On the other hand, each one of [239], [246], [256], [258], [259] is capable of adapting to different workload types without re-training. All predictors here are event counters. However, events are not necessarily linear predictors of power consumption across all workload types. In [261], it was found that the model parameters of a linear regression of event counters onto power consumption are workload specific. Given the contrast with [73], [74], it seems that the root cause is the selection of events for prediction.

Concurrent operation: In [238], a non-linear model of the dynamic power of a multi-(virtual)-core VM is obtained. The dynamic power p_{vm} is expressed in terms of the average utilization $u_v$which n virtual cores impose on a total of N physical cores:

$$p_{vm}(u_v, n) = \alpha \left(\frac{n}{N}\right)^{\beta} u_v^{\beta}$$

Parameters α and β are determined through linear regression of the logarithmic form. The relationship was successfully tested under the operating context of one VM and three VMs. The model is limited to processor-intensive workloads; indeed, this is not surprising, since the predictor is (virtual) processor utilization. However, a conclusion can be drawn: the accuracy of the model in predicting the dynamic power of an individual VM is independent of concurrent operation of other VMs (albeit for a limited range of workload types).

In [252], it is shown that estimation based on processor utilization can accurately predict the power consumption of multiple concurrent VCs. The prerequisite is that a model for that VC's workload type has been learnt. The prediction accuracy for individual VCs is independent of concurrent operation. The same conclusion can be drawn for [238]; here, furthermore, the range of workload types has expanded to a broader range. With regard to results obtained in [261], the conclusion is yet again the same: individual prediction accuracy (albeit workload-type-dependent) is independent of concurrent operation.

Variability in both workload type and number of concurrent, co-hosted VCs is manifestly claimed in [239], [246], [258]. It is noteworthy, therefore, that not only is th203odellingng system adaptable, but it also produces a model that is itself adaptable without the need for real-time adjustment.

I conclude this sub-sub-section with the observation that limited adaptability (defined as independence of workload type and number of concurrent, co-hosted VCs) has been achieved. From surveyed RUs, it can be concluded that presently, the limits of adaptability are the following.

1. Workload type: Power consumption of processor-intensive, memory-intensive, disk-intensive workloads and mixes thereof, has been modelled by a single modelling system in an automated manner.
2. Co-hosted, concurrent VCs have been modelled up to but not exceeding over-commitment of processor physical cores.

Caution has been exercised in claiming these limits. For example, while commitment of logical cores (e.g., Intel Hyper-Threading logical cores) has been investigated, no evidence has been found that power consumption has bee203odellined in a manner that automatically adapts to a transition of consumption from physical to logical cores.

3) Lack of use of metrics of energy efficiency and standards to address this deficiency

I note several experiments, e.g., [232], [233], [245], [248], [280], [281], [282] that target power consumption but less than 12% of the RUs in the corpus approach the problem in terms of some energy efficiency metric [226], [227], [242], [262]. It is necessary to move beyond measurements of how much power was consumed, to measurements of how much power was consumed to carry out a specific task. This change in approach facilitates comparison between research works. More importantly, it directly addresses the question about cost of operation of infrastructure.

This approach requires identification of a unit of comparison that transcends the boundaries of disparate systems that deliver this unit. This unit of comparison is referred to in the Life Cycle

Assessment (LCA) framework (ISO 14040) [91] as the ***functional unit***. A definition specific to telecommunications equipment is given in [283] : the functional unit is defined as "a performance representation of the system under analysis". This definition is too broad; therefore, units specific to a variety of classes of equipment are defined too [283]. In the corpus, two approaches that have been seen are hash/J [227] and J/Web Interaction [242]. The functional units in these cases are performance of one crypto-hash and one web interaction, respectively. Another is to define a functional unit specific to a digital service delivered over a telecommunications network, e.g., ten minutes' time of browsing [94]. Note that L.1310 [283, p. 4] recognizes both metrics where energy is in the numerator [242] as well as those where it is in the denominator [227]. Guidance is available: energy efficiency measurement for several NFV components has been standardized, as well as measurement standards for servers, switches and virtualization systems [284].

4) Trends

Here, I present trends which have been identified through collation of problems, approaches and developments. This sub-sub-section is divided into four parts, regarding trends in: challenges, complexity of tackled problems, approaches and developments. Within each part, trends are numbered using Arabic numerals, to differentiate cleanly between them.

Challenges

1. The causality DAG (Fig. 58) shows that the research space can be characterized succinctly: note the thickness of the links originating at P7 (VC resource use and measurement) and P8 (how to attribute host system power to VCs).
    a. Fig. 41 and Fig. 60 show that the challenge-category tackled most frequently ($R_{P_7} = 25.3\%$) is how to load the VC's resources and measure the loading (P7). P7 also has the most frequent presence in RUs: $F_{P_7} = 65.7\%$) (Fig. 60).
    b. The accurate specification and measurement of load is essential to model formation. These measurements provide the aggregated (i.e., indiscriminate of which VC is consuming) predictors – the input side of the model. Disaggregating the predictors and attributing measured power consumption to the individual VCs (the output side of the model) constitutes the second most frequently-tackled challenge ($R_{P_8} = 16.5\%$). P8 also has the second most frequent presence in RUs: $F_{P_8} = 42.9\%$.
    c. In the following, the vector (P, A, D, weight) will be used to indicate a path (P, A, D) and the number of triads (weight) through the path.

   i. 34.2% of triads collected regard P7 and P8 respectively. The triads graphic (Figure 19) shows that efforts rooted in these two challenges converge on a common objective: building adaptable models of VCs' power consumption, notably using regressions to linear combinations of computing resource predictors.
   ii. The set of triads leading to linear regressions consists of {(7,10,5,20), (8,17,5,13), (7,9,5,12), (8,1,5,10), (8,2,5,8), (7,8,5,6), (8,12,5,4), (7,6,5,3)} and these account for 11.3%.

2. The third most frequently-tackled challenge is that of estimation of virtualization-host-system power consumption ($R_{P_8} = 15.4\%$). This category might be overlooked in a first inspection of the research space, as it might seem futile to attempt to estimate a power consumption which can be measured. However, in practice, the logistical challenge of measurement of horizontally-scaled system deployments seems to be well known and several works have been undertaken to develop software meters.
3. Several significant links originate on P1 (profiling power consumption's dependency on architecture). The DAG (Figure 18) indicates that approaches to tackling P1 are spread across a mixture of managing resource provision, use of synthetic (resource-specific) workloads and prediction using system software's instrumentation
4. Another large group of links originates on P3 (estimation of host system power consumption). The DAG (Figure 18) indicates that a primary concern in tackling P3 is the type of model to select. The most common choices are linear and non-linear; machine-learning techniques are the least common of the three. The thickness of the triad (3,2,10,11) (D10: relative accuracy of formal approaches), indicates that there is already significant interest in whether the advanced models are worth the effort to develop them and computational resources required to operate them
5. While $F_{P_7} = 65.7\%$ (loading VCs and measuring their use of resources), only 8.6% ($F_{P_{7.2}} = 8.6\%$) of all RUs investigate loading containers and measuring their use of resources. With virtual data plane devices, the figure is even lower: $F_{P_{7.3}} = 8.6\%$. This imbalance suggests that there is much room left for research into modelling power consumption by containers and data plane devices

Challenge complexity

In [223, p. 13], I suggest that diversity of approaches is a metric of the complexity of a challenge. It can be seen (Fig. 60) that challenge complexity (W_{P_k}) generally follows the frequency with which a particular challenge-category is tackled. That is: the more frequently the challenge is addressed, the more diverse are the approaches applied to it. This can be verified by noting that the heights of both sets of bars in the chart (superimposed on the same graphic) follow roughly the same pattern. However, some categories do stand out. For example:

1. P6 regards the influence of temperature on the power model. Approach diversity is poor here because only the use of additional instrumentation can be attributed to this challenge. Model bias might be attributed to this challenge too but largely, model bias is determined by other challenges within the scope of the RU.
2. While host system power attribution (P8) has the second highest research interest (and frequency of occurrence, F_{P_k}), the number of approaches taken to solve this challenge is relatively small
3. On the other hand, power consumption's dependencies are tackled by a disproportionately large number of approaches. This is not surprising, as the objects of study (architecture, virtualization platform, virtualization genre) are multi-faceted and dependencies can be investigated through a variety of approaches

Approaches

The approach utility metric, U_{A_k} [223, p. 13] (Fig. 61), seems to be a useful one. It communicates clearly what has been observed during surveying. Below, I draw attention to saliencies perceived during surveying and confirmed by the metrics.

1. Instrumentation of consumption of computing resources (A10 – which includes microarchitectural instrumentation and that abstracted by system software) is repeatedly adopted ($U_{A_{10}} = 15.5\%$, $R_{A_{10}} = 15.7\%$) in empirical work in this field. It is also the most utilized of all approaches. In comparison, workload instrumentation accounts for 8.2% utilization ($R_{A_{11}} = 5.2\%$).
2. Resource-specific workloads (A8) are the more utilized approach to loading VCs (14.3% of all triads). This approach category is the second most utilized. Workloads representative of real use (A9) account for 9.0% of all triads. The corresponding frequency of occurrence figures (R_{A_k}) are 13.4% and 10.0% respectively.
3. Two other high-utility approaches are (a206odellingng bias (A2, with $U_{A_2} = 11.9\%, R_{A_2} = 11.6\%$) and (b) the managed provision of resources (A6, with $U_{A_6} =$

10.2%, $R_{A_6} = 10.8\%$). Use of representative workloads (A9) follows at $U_{A_9} = 9\%, R_{A_9} = 10.1\%$.

4. A comparison of the patterns of bar height distribution for U_{A_k} and R_{A_k} reveals that some categories stand out.
    a. While A12 (direct power measurement) is not as frequent as the other software-based forms of instrumentation, this approach has a utility that sticks out of the pattern (Figure 21). The reason is that most, if not all, developments obtained in an RU that includes this approach depend on the directly measured power.
    b. Similarly, while workload characterization and instrumentation (A11) is employed with a frequency that is about one-third that of its alternative (i.e. computing resource instrumentation, A10), it has a far better utility-to-frequency ratio than A10. The triads graphic (Figure 19) indicates that one important cause of this high utility is that software-L2-data-plane switches and (virtual) network functions are investigated primarily using measurement of workload (and not measurement of computing resources consumed to process a workload).

Developments

1. 55.6% of all identified developments are obtained in modelling power consumption. The remaining 44.4% regard how power consumption depends on implementations. Implementations investigated (for their impact on power consumption) range from entire virtualization platforms (e.g., KVM) to components (e.g., processors).
2. The most frequent ($R_{D_{12}} = 14.5\%$) category of developments is that which regards observations on dependency of power consumption on architecture. The cause of this high frequency is that developments in the study of architecture establish directional (negative, positive, neutral) correlations rather than predictive forms. For example, in a single RU [244], all of the following developments emerge:
    a. D12.08: VM power consumption increases linearly with vCPU frequency of operation when the vCPU is operating at 100% utilization;
    b. D12.09: Virtualization-host-system power consumption increases linearly with the number of physical cores operating at 100% utilization;
    c. D12.1: Virtualization-host-system power consumption increases with the number of VMs active on the same core
3. Amongst developments of models, the most common category (D8) is of the type where power consumption (host or VC) is predicted in terms of workload characteristics ($R_{D_8} = 12.0\%$). The next most common category regards prediction of

power consumption as a linear function of computing resources ($R_{D_5} = 8.5\%$)). Machine-learnt models of VC power consumption comprise the second least frequent category, with $R_{D_7} = 1.7\%$.

4. Despite the frequency of developments in category D10 (Figure 22), few works [237], [239], [240], [254], [256] compare accuracy of model types. Models have largely been treated as a means to an end, with little investigation of their relative accuracy and range of validity (in the seven dimensions of variability). This may be a reflection of researchers' interests. As the popularity of methods from the body of knowledge of data science increases, works (e.g., [240]) that span a broader range of model types may be expected to increase in concert.
5. Adaptable, non-linear VC models (D2) occupy a modest 4.3%. Such a distribution adds weight to the observation that this research space is ripe for exploration using advance208odellingng techniques. Indeed, one attraction for data scientists is the relative ease with which data can be collected. However, polynomial and other types of regression to closed-form are problematic as suspicion of over-fitting increases with the order of the polynomial. For example, in [253], a sixth-order polynomial is suggested to model the relationship between processor utilization and power consumption by a host.

5) Three levels of abstraction

It was noted that existing power models may be classified into one of three levels of abstraction. In ascending order of abstraction, these are:

- Microarchitecture and architecture
- Simple characterization of workload
- Complex characterization of workload

The meaning of "abstraction" specific to use here is perhaps most easily grasped by referring to the variables used as model inputs. In all cases, the variables are some measure of load. At the lowest level, inputs that quantify operation of processor sub-units are used (event counters and event timers). The highest level uses inputs that quantify the demand for a telecommunications system or service. Clearly, the latter inputs are far more detached from the underlying, concrete implementation than the former.

With regard to the lowest level of abstraction, many power models use hardware resource consumption to estimate the power consumed by virtual components. One survey [282] (while comparing the available power models for processors, VMs and servers) observes most of the power models for virtual machines use physical machine counters to estimate the corresponding resource utilization by the virtual components. Further insight can be offered on this matter. I concur in the observation that much current research is concerned wit209odellingng power consumption of virtual machines. The approach may succinctly be described as estimates obtained from models trained out of either architectural or microarchitectural instrumentation data. Note that I distinguish between architecture and microarchitecture using the classical interpretation [285, p. 17]. Now, the term "architecture" is severely overloaded and its interpretation can easily differ from that which I wish to use. In the following description of the levels of abstraction, the alternative "system software's instrumentation" is used to convey the same meaning as "architectural instrumentation", with less ambiguity.

Microarchitectural instrumentation is the lowest level of instrumentation. Power consumption is expressed in terms of variables that are defined at sub-CPU and sub-subsystem levels. The granularity of this level holds the greatest potential for accuracy, but the rate of change of observed variables has deterred several researchers from pursuing this approach to instrumentation, citing concerns about communicational and computational overhead. This concern has been dismissed by three groups of researchers [247], [263], [275], who have indicated that acceptable accuracy can be achieved with negligible overheads. Given the proliferation of works based on this approach, the availability of high-level language code that facilitates use and the potential for capture of physical behaviors, then ***my general recommendation is a preference to investigate use of microarchitecture instrumentation***.

System software's instrumentation regards a class of instrumentation that has meaning across the spectrum of computer systems. The input variables, such as CPU or network interface utilization, are produced by some digest (function) of intermediary system software. For example, an operating system's (OS) measurement of core utilization can roughly be described as the core's duty cycle on behalf of the OS. This is significantly removed from knowledge of activity within the core. While models at this level are more abstract, they are still low-level, especially when compared with the other levels.

Simple characterization of workload is a less frequently encountered abstraction, used by Enokido, Takizawa and various others with whom they have co-published [245], [248], [286]. These models describe power consumption in terms of fundamental descriptors of workload, e.g., number of processes and transmit/receive data rate.

Complex characterization of workload is the least granular of the models in the survey [82], [287]. The objective here is to quickly proceed to a good estimate of the power or energy required to produce the workload. This kind of model has no use in real-time control ***but it is useful for macroscopic comparisons***, i.e., comparisons between two disparate systems for provision of a service. The comparison might regard two different paradigms of provision of the service, e.g., classical vs virtualized implementations. Thus in [82], the implied (system) metric is the amount of power required to deliver 1 million packets per second of throughput through an evolved packet-core's (EPC) serving gateway (SGW). In [287], [288], [289], the objective is to minimize the amount of consumed power by virtualizing baseband processing functions, evolved packet core, customer premises equipment, and radio access network functions.

6) Service determinism: a criterion particular to the telco cloud

The NFV data plane demands determinism [85, N. see video @25:45], [273], [290]. Strictly, ***service determinism*** is sought, since the packet arrival process is generally unconstrained. The root cause of this need is to correctly size equipment resources to meet load demands, whether throughput, latency or jitter. ***This need is intense***, as it impinges on a PTNO's obligations, specified in legally-binding service-level agreements (SLAs). Service determinism has been approached through augmentation of GP hardware, with domain-specific architectures (DSAs, which was first referred to while describing problem category P1). Hardware-centric DSAs for the NFV data plane are constructed out of elements (or systems thereof) that can be divided into three groups.

1. **IO system architecture:** This group comprises the abstraction, ***directly at a peripheral interface***, of functionality that facilitates virtualization of hardware, e.g., SR-IOV (used in PC– - Peripheral Component Interconnect) and N-port ID virtualization (Fibre Channel).
2. **Processor architecture:** This comprises architectural change that facilitates partitioning of processor resources, such as multi-core processors and NUMA.
3. **Co-processing:** Compression and decompression, encryption and decryption, and packet header processing are examples of high-volume tasks that can be offloaded to co-processing subsystems. Examples include Intel QuickAssist Technology (QAT) and TCP Offload Engine (TOE).

A further a set of approaches is observed that are complementary to DSAs in the quest for determinism. These include:

1. **Large memory pages:** exploitation of the facility to organize virtual memory into larger pages than the general-purpose 4 KiB;
2. **Userspace programming:** diversion of control of hardware resources away from the multi-service kernel, to single-common-use user space programs, e.g., DPDK and AF_XDP.

Notably, resource partitioning (see processor architecture, above, i.e., the second hardware-centric approach), combined with userspace programming (the second complementary approach), realize the run-to-completion scheduling model [273].

I conclude with an observation on ***the cost of current realizations of service determinism***. The (aforementioned) combination of core partitioning and userspace programming has been widely adopted through Intel's popularization of DPDK, via Intel's open-source liaison efforts. The multi-core processor enables service determinism through an approach that is anathema to the principles of multiprogramming: dedication of hardware to a specific task.

7) Direct or indirect measurement of power in virtualized environments?

It was observed that most research in modelling power consumption seeks to obviate the need for direct measurement through indirect measurement. This indirection consists of measurement of resource use which has a discoverable relationship with power consumption by the entity hosting the resources. Modelling, here, has the objective of indirect measurement of a variable that is not directly accessible (power consumption by VCs), through others which have convenient and reliable instrumentation. The accessible variables are referred to as ***power proxies***. The RAPL interface provides a unique approach to measurement aI it directly addresses power consumption. However, notwithstanding appearances of direct measurement, RAPL is actually based on a software model that uses performance-monitoring counters (PMCs) as predictor variables to measure power consumption. It is available in processors starting from the Sandy Bridge microarchitecture. RAPL measures the power consumption of different physical domains, where each domain consists of either cores, sockets, caches, or GPU. I briefly comment on its accuracy through references to research that has investigated them.

1. In [275] the advantages and drawbacks of using RAPL were investigated. Different Intel architectures such as Sandy Bridge, Haswell and Skylake were used in the experiments to analyze RAPL's accuracy and overhead. Data collected were modelled using a linear model and a Generalized Additive Model (GAM). Accuracy of predicted results was compared with the measured power consumption from a precise external

hardware power meter where RAPL based models show 1.8 – 4.3 % of error for the various architectures. Prediction accuracy of RAPL-based power models was also compared with those based on OS counters, where OS-based models show a higher error of 5-16%. Also, the performance overhead (in terms of time) of using RAPL was studied at different sampling frequencies and for different application runs. Results show that even with high sampling frequency of 1100 Hz, RAPL incurs overhead of not more than 2%. Some limitations of using RAPL include: poor driver support to read energy counters, overflow of registers due to their 32-bit size and measurement of energy consumed by individual cores.

- Another study to analyze the precision of RAPL is presented in [291], where only the dynamic change in power consumption is observed. An external power measurement unit (WattsUp Pro), is used as a reference for power measurement values. Intel Haswell and Skylake servers were used in the experiments to run different applications and to find the reliability of RAPL with the help of external power meters. However, in this research work, only two power domain packages (power consumption of whole socket and DRAM domain of RAPL) were observed. Applications such as dense matrix multiplication and 2D Fast Fourier Transform were used for server power profiling. Results show that the power measurement error varies with changing application and its workload size. For different applications the average measurement error using RAPL was in the range of 13-73% considering WattsUp power meter as the ground truth. It was concluded that with the modern multi-core parallel processing and resource contention for shared resources, there is a complex non-linear relation between performance, workload size and energy consumption. Hence it is difficult to attain low error percentage for power measurement using on-chip sensor

7.4.2 Research gaps

Three significant challenges remain unaddressed, while a fourth requires further attention:

1. modelling of containers power consumption;
2. the effect of overcommitment on power efficiency;
3. investigation and classification of DPDK applications and

4. the fourth challenge, which is starting to receive some attention [260], [262], [264], [271], regards modelling of power consumption by virtualized I/O.

The following treatment of research gaps does not address improvements in approaches. It has already been indicated (in the treatment of developments) that more adaptable modelling methods are required to tackle the dimensionality of the field. Similarly, lack of use of metrics of energy efficiency are not included with the research gaps, as it is a deficiency in the approaches, not a challenge in itself. Rather, here, attention is focused on where the more pressing challenges lie for development of power and energy control of VCs.

Gap #1: Modelling of containers' power consumption. Few works [255], [258], [259] tackle containers from the perspective of their power consumption. Yet, at least for the telco cloud, VMs are no longer the destination (see, for example, [88] and [292, N. see video @10:30]. Containers have replaced virtual machines as the base for deployment of virtualized network functions. In [255], the approach(-set) taken is to:

1. use representative workloads e.g., HEVC (High Efficiency Video Coding) transcoding and machine learning image classification, and
2. customization of the set of low-level instruments used to correlate power and energy consumption with workload characteristic.

This work develops a profiling tool. It provides guidance that is specific to application and both the hardware and software aspects of the containing platform. "[P]olicy" for "tradeoff between energy, power and application performance" is the cited objective. Given the high-dimensionality of the core challenge, this approach-set to modelling and measurement of power consumption may well be more coherent with the European Telecommunications Standards Institute's (ETSI) Management and Orchestration (MANO) standard. Such information would then be included in the infrastructure-resource-requirements meta-data descriptors in the VNF package [293].

Gap #2: The effect of overcommitment on power efficiency. Overcommitment consists of the allocation of more capacity of some compute resource to VCs, than is physically installed. The concept is very similar to oversubscription of telecommunications capacity to subscribers, such as when the arithmetic sum of capacities of access links exceeds the aggregating device's backhauling capacity to a central office/local exchange. In this context, overcommitment principally concerns processing cycles and memory space. As with oversubscription, there is an optimization problem to solve. One problem of interest to this survey's scope is understanding the relationship (say, ratio) of committed virtual resources to installed physical resources that optimizes total cost of ownership (TCO) of cloud infrastructure:

- on the one hand, the facility to overcommit has a direct impact on the density of packing of VCs (number of concurrently active VCs) on a virtualization-host-system, thereby reducing the TCO;
- on the other hand, overcommitment may reduce the power efficiency of a workload.

Had this challenge been tackled in any depth, or at least in any breadth, it would have merited a category of its own. Currently, however, I am only aware of a single study [242] that tackled this challenge. Results obtained strongly justify overcommitment of processor cores to vCPUs, for the case of transaction web service workloads, with the increase in throughput (measured in web interactions per second, or WIPS) increasing at a faster rate than power consumption. This behaviour was observed well into overcommitment ratios of processor cores to vCPUs equal to 3 (three). Overcommitment of physical to VM memory was not investigated.

Gap #3: Investigation and classification of DPDK applications. In a previous sub-sub-section, the relationship between DPDK and power consumption is addressed. In the course of a public discussion in the forum offered by the North American Network Operators Group (NANOG) [294], there emerged a need for clarity on DPDK's association with inefficiency in power consumption. Interest was particularly expressed in knowledge of a classification of extant DPDK applications according to their power consumption, and contribution to the code base to improve power-hungry applications.

Gap #4: Modelling of power consumption by virtualized I/O. Power consumption of network I/O has been investigated to some extent as this is central to the feasibility of network functions decoupled from hardware. Yet while software and hardware solutions are already available they require frequency and idling control targeted to their specific operating conditions. Notably, naïve DPDK runs the processor core at its maximum power consumption, regardless of load. Exploitation of adaptive-rate (AR) processing and low-power idle (LPI) should provide a means to save power while processing high networking loads. However, effective control of these means is still elusive, despite both using Xeon Haswell microarchitectures, [260] and [264] reach opposite conclusions about the feasibility of processor core C-states. The former [260] finds LPI an effective means of reducing power consumption of packet forwarding (with limitation on latency) while the latter [264] finds it ineffective, preferring use of the **pause** instruction. Furthermore: in [277], performance state transitions (P-state) are found to impose a high transition latency, while in [264], P-state regulation is the preferred approach. There is scope for research in the dynamic adaptation of the processor's operating state to save power.

7.4.3 Pitfalls

1) Power consumption does not in general increase linearly with processor utilization

Notwithstanding advances made in identifying operating contexts that manifest sub-linear power-utilization relationship [239], [246], more recent publications [234], [295], [296], [297], [298], [299], [300], [301], [302] persist in using the linear model without acknowledging its limitations. The model is simple to use and has some foundations in research [303]. It has three premises, described here with regard to the operation of Microsoft Windows:

1. When Windows has no threads to run on a logical core, it schedules the idle thread [304].
2. The idle thread keeps the processor in a low-power state [305]. The specific state depends on the processor's green capabilities.
3. In the complement (non-idle time), the processor issues instructions at a constant rate.

This simple model has limitations [239, p. 808], [306, p. 6]. It fails to take into account diverse processor operating contexts, some of which are coming to bear on current use cases. Specifically, the third premise is true only to the extent to which instructions are being fetched and data are being loaded from/stored to instruction and data cache, respectively. Consider the context of 90% and greater hit ratios. At such cache hit ratios, the rate of instruction issue is expected to be narrowly distributed about its mean. By contrast: the lower the hit ratio at the cache level before main memory, the lower the fraction of non-idle time at which power consumption saturates. This saturation is strikingly illustrated in [239, Fig. 1]. Variation of power consumption due to execution of tests from the SPEC CPU2000 benchmark suite is shown. The power consumption diverges at 25% CPU utilization and the consumption of the processor-bound test (mesa) is greater than that of the memory-bound test (mcf) ***by a factor of about 2.6***.

Another good (albeit broad) illustration of this pitfall is given in [307, Fig. 5]. Data on power consumption and CPU utilization under a standardized benchmark is plotted for four different physical server models. None of the relationships is linear. Neither is there a single, common behaviour.

Researchers align themselves into two groups in regard to CPU utilization. One group favours (operating-)system metrics (of which CPU utilization is one metric) and the other favours event counters (microarchitectural instrumentation). The arguments posed by each group against the other's approach can be summarized as follows. The "system metrics" group claims that the "event counters" group's work is (a) not portable (at least across microarchitecture families) and (b) cannot be exercised without low-level access to the host (thereby, this approach cannot be exploited by user-level privileges) (see, e.g., [308, p. 121] and [252, p. 43]. The "event counters" group claims that CPU utilization is a workload-dependent predictor (see, e.g., [236, p. 1380]) and therefore cannot be used without re-training the model. Indeed, this modification to the "system-counters" approach is employed in [253, Sec. 3.2]), where it is stated that "[b]ecause of changes of VM's internal applications … parameters must [be] recalculated automatically". ***Given these arguments, it seems that the system metrics group argument is weak: both system metrics and event counters require re-training if hardware is changed but system metrics lack the granularity to discriminate between workloads*** (cf. [239, Fig. 1]). This means that CPU utilization can only be used as the sole predictor if it is re-trained with change in workload. This problem – which I have termed the fallacy of the universal model – is dealt with, in a sub-sub-section of the thematic analysis.

As regards use of hardware threads (Intel® Hyper-Threading), I have observed that various works concur on the operating context under which a linear relationship is subject to the lowest error. This includes at least the following two conditions.

1. The processor cores are increasing their instruction issue rate in proportion to the fraction of time they spend busy. This implies that instruction and data cache hit ratios are high. This is simply the third premise.
2. Only one logical core is active per physical core at any given time [239], [240], [246]. Expressed alternatively, actual utilization must lie below half maximum utilization. The underlying cause is that activation of the second logical core employs fewer organizational units of the processor than activation of the first logical core.

The first condition is particularly problematic, as cache miss ratios are likely to be much higher in the context of virtualized environments. In such environments, the number of runnable threads is the sum of runnable threads controlled by independent operating systems. Evidently, this is higher than the expected number of runnable threads on a single server instance.

Other evidence of this "utilization trap" is not hard to find. In [234], the compute resource is stressed using cpulimit and stress-ng. The "cpulimit" utility runs a specified process image, then pauses and resumes it until a certain percentage utilization is reached [301]. The repetitive execution of a single process is highly likely to create conditions for very high instruction- and data-cache hit ratios. Such favourable hit ratios skew results towards the linear relationship between CPU utilization and power consumption.

2) DPDK is not intrinsically inefficient in power consumption

Research on power efficiency in DPDK applications [225], [228], [230], [264] has portrayed DPDK as power inefficient. Before proceeding to an exposition of this pitfall, it is necessary to distinguish between data, control and management planes. “Data plane” is a term used to refer to the infrastructural means that provide the capacity for exchange of customer (or subscriber, or end-user) data. It is complemented by a control plane, which refers to those means that facilitate the dynamic setup, maintenance and tear-down of a functional data plane. Another complementary part is the management plane. This includes the infrastructural means for a network operator to configure and monitor the control plane and the data plane, as well as intervene to correct faults arising in either plane. Simpler networks may have no control plane.

Now, I proceed to the exposition of the pitfall. In one particular case [225, p. 43], it is claimed that "we found that a poll mode driver (PMD) thread accounted for approximately 99.7 percent CPU occupancy (a full core utilization)." The implication that seems to emerge here is that the PMD itself is driving this power consumption.

This portrayal is problematic at best and incorrect at worst. The referenced investigations of DPDK have indicated a very low power efficiency, **but they do not clearly distinguish between responsibility of the DPDK API and the application using it** (the API). A recent, public thread [294] has emphasized the responsibility of the application developer in the avoidance of the naïve, "default approach" of busy polling. Such an approach would, indeed, poll network IO hardware continuously [225, p. 44], truly fitting the epithet "spinning-hot" [85, N. see video @26:23]. However, a broader (in ”he sense of including industrial correspondents) investigation [294] suggests that:

- contrary to claims in [225], it is the driving behavior of OvS that is inefficient in power consumption, and
- there are simpler, technical means of throttling a polling loop, including, say, the use of program code to interleave ACPI C1 states with polls according to traffic demands.

These latter observations cast doubt on the claim that automated frequency control is outside the scope of current frequency governors, since "the OS won't be able to distinguish whether it's under a heavy load" [225, p. 45]. On the other hand, savings through NUMA awareness [228], (where transmit/receive port, memory and processor core are kept within the same NUMA node), is affirmed in [294].

3) Research on power models without power-relevant context

This pitfall traps readers who attempt to draw conclusions from published research which lacks a clear specification of context relevant to power consumption. The pitfall is best illustrated through examples.

1. **Failure to emphasize context: idle power consumption vs frequency**. Dependence of idle power consumption on clock frequency is context-sensitive. In [226], it is explicitly stated "idle power consumption remains constant, regardless of the CPU frequency … across the whole frequency range" (1.6 – 2.6GHz). The CPU is an Intel Xeon E5620. In [258] and [259], a quadratic relationship between idle power consumption and frequency is observed. Here, the CPU is an Intel Core i5 Haswell. In these two instances, emphasizing the restricted scope of findings would suffice to spare a reader from excessively broad inferences.
2. **Failure to emphasize context: idle power consumption vs hardware and software specification**. Enokido's and Takizawa's work [248] derives a power consumption model for a server while VMs run computation-bound processes. The servers used run on Intel Core i5-3230M processors. These processors are used in the mobile device market [309]. They are capable of low-power idle states [310]. CentOS 6.5 uses a tickless kernel [311]. Combined, these facts, relevant to the context of power consumption, provide a plausible explanation for the observed increment in power (denoted, in [248], by $\min C_t$), when a core in a package is activated. Again, therefore, the scope of findings is likely to be restricted
3. **Failure to fully define context: Configuration of power-relevant parameters.** I use [312] as an example. No reference is made to whether Hyper-Threading is enabled. This is essential to understanding how the ESXi vCPUs are created. Neither is any information given about how the vCPUs are related to physical (or logical) cores. Nor is the reader told how virtual network interfaces and switching are set up. ESXi version 5 offers both paravirtualization ("vmxnet") and emulation ("e1000") to implement virtual network interfaces. The impact on energy consumption of selecting a virtual network interface implemented by emulation can be expected to be high [233].

The examples cited illustrate the importance for a researcher into power models to qualify his/her results ***with a well-defined physical context***. Research into power models involves hard components and a diligent characterization thereof is essential to the acceptance of work as scientific research.

4) Benchmarks may skew power consumption according to their organizational dependencies

It has been shown that both "cpulimit" and "stress-ng" do not produce generally representative measurement of power consumption. This observation is not limited to measurement of power consumption. Use of kernels, toy programs and synthetic benchmarks to measure performance has been identified as unrepresentative [285, p. 40] of general performance. Benchmarks are standardized workload generators that are used for comparison of computer systems for a specific class of application. Unless this application class is a good representative of the application of the computer system in productive use, the power consumption measured under test is not a reliable predictor of that obtained during productive use. It is necessary to plan test workload generators in advance and state the limits of validity of results. In [242], TPC-W is used, which is a transactional web benchmark that can simulate the business oriented online web-servers. The MySQL++ Java version of TPC-W benchmark, suitable for cloud applications, is used to generate the online traffic, where three different traffic profiles based on browsing, purchasing and ordering of books are generated. The throughput measure for these servers is observed through the metric Web Interactions Per Second (WIPS).

5) Processor organization significantly impacts power consumption

This point is illustrated with a wide-ranging example [313] which compares the Intel Xeon X5670 and AMD Opteron 2435.

1. Different idle loops (using no operation, pause, repetition, etc.) were tested to see their effect on power consumption of both systems. It was observed that the Intel Xeon has a loop stream detector, which disables the processor's features like fetch and decode. On the other hand, the AMD processor has no hint to process these loops efficiently; hence, it consumed more power than the Intel processor.
2. A processor consumes a different amount of power depending upon the instruction (such as load, addition, multiplication, etc.) and the level in the memory hierarchy which is accessed by the instruction.
    a. For the Xeon, data throughput of all instructions from a particular memory hierarchy level is almost the same, but there is a difference in their power consumption. The 'load' operation consumes the lowest power compared to other instructions, and this holds true for all memory hierarchy levels. The reason is that the 'load' instruction just needs to load the content on the processor registers whereas 'add' and 'mul' operations are more computation demanding.
    b. However, the AMD processor's behaviour is the opposite. When the 'load' operation accesses L1 cache, it achieves almost one-and-a-half times the data throughput of other operations and hence also consumes more power. This difference in resource utilization is due to the different microarchitecture of AMD processors, where the 'load' instruction is handled by many floating-point pipelines. Other instructions just use a single pipeline for their operations. Moreover, AMD processors have an exclusive cache level design, which requires write-back when evicting data among different cache levels. Since

Intel’s inclusive cache design does not require this function, it consumes less power. Within higher memory hierarchy levels (L2 or L3 or main memory), the AMD’s computation (‘add’ and ‘mul’) and data transfer operations (‘load’) deliver roughly the same data throughput and consume roughly the same power.

6) Isolation of VC for power modelling and measurement

Isolation of any VC from its hardware counterparts cannot be done completely [314]; thus, the assumption of measuring power consumption of an individual virtual entity irrespective of the hardware on which it is implemented is an illusion. The virtual infrastructure is composed of several components at both hardware and software level, where the effect of underlying hardware, OS and VNF technology can significantly impact the power consumption. Hence, isolation as well as modelling of power consumption for an individual virtual component is difficult to obtain.

7.4.4 Fallacies

1) A universal power model

I have suggested that the core challenge in modeling power consumption by VCs lies in the number of dimensions of variability. This has been demonstrated throughout this survey, where a number of generalizations have been addressed. Summarizing, the literature shows that:

1. host power consumption does not generally have a linear relationship with processor utilization;
2. CPU-intensive workloads that repeatedly execute the same code skew power consumption results;
3. network-intensive workloads are power- and time-consuming because they employ emulations of network switches, but the root cause (emulation in the hypervisor software switch) disappears with SR-IOV [315, p. 5];
4. host saturation must be taken into account in predicting VCs' power consumption;
5. processor utilization (an architectural attribute) is insufficient to predict host power consumption and microarchitectural attributes, such as LLC misses, are necessary to predict host power consumption even for the same level of processor utilization.

This list, while not exhaustive, amply illustrates that the several dimensions of variability are significant in determination of VC power consumption. A model claiming to determine power consumption as a function of fewer variables than the dimensions that have been pointed out, ***must be accompanied by a scoping region that limits its use***. While a precise scope may be an unrealistic demand, it is essential that guidance be given about conditions of use of the model. I now illustrate this point by using two examples from the corpus.

Example #1: Khan [227, p. 51] compares energy efficiency (hash/J) obtained by scheduling process threads on additional cores, with that obtained by scheduling them on hardware threads on active cores (through Intel Hyper-Threading). He shows that the former is greater than the latter. In apparent contrast, Enokido and Takizawa [245, p. 279] show that for a given data transmission rate through the uplink of a software virtual switch, greater energy efficiency (W/bps) is obtained by operating an additional hardware thread on an active core (through Intel Hyper-Threading), than operating an otherwise idle core. An important difference lies in the task's processing "intensity", i.e., the rate of supply of instructions. While Khan's operations are tightly bound to the processor (cryptographic hashing), Enokido's and Takizawa's operations are distributed over the processor and network input/output. Without delving into detail, it is realistic to hypothesize that the average instructions per second demanded are far lower in the networking application, since transmission of a large file (as is the case here) does not take place in one processing burst. Operating time is divided between the processor and the media channel. In such a scenario, the added capacity of the same-core hardware thread suffices.

Example #2: At the time of writing, the scope of validity (where the scope is a sub-space of the seven-dimensional space) is typically only implicit. Notably, in [252], a "refined model" is used as a means of accurate prediction of power consumption by virtual machines while running very specific benchmarks. It is also noteworthy that the authors contemplate a type of onboarding process wherein "new" VM entrants to a cloud are modelled as a prerequisite to their inclusion in the power-prediction system. Indeed, such a process is already intrinsic to management and orchestration of virtualized network functions. Just as the virtual deployment unit (VDU) nodes (in virtualized network function descriptors (VNFDs)) store VM properties describing computer system resource demands, so can the descriptor template be extended to provide properties regarding power consumption demands. This "onboarding" is necessary since the selected predictors and modelling do not cover a sufficiently broad range of workload types, and a specific model must be learnt online, i.e. – on the fly.

On the other hand, a comprehensive power model for existing implementations may be possible, notably when the following two conditions hold true:

1. Every resource that consumes power must own a counter that registers its usage, or lack thereof, during a specific clock cycle.

2. Usage of a specific resource during a specific clock cycle must consume a constant amount of energy. This has the following corollaries.
    a. Energy consumption by the specific resource is a linear function of the number of clock cycles for which the resource is active.
    b. Power consumption of a system can be expressed as a linear combination of the total set of such resources.
    c. The amount of energy consumption by a specific resource during a specific clock cycle must be independent of usage of other resources during any other clock cycle.

7.5 Yet another recap

This section summarizes the chapter's contributions (sub-section 7.5.1) and suggests a framework for future research into real-time, predictive models of power consumption by VCs (sub-sectio2).

7.5.1 Contributions

Seven dimensions of variability have been identified (workload type; virtualization agent; host resources and architecture; temperature; power attribution; co-hosted, concurrent VEs and (clock) frequency of operation) and observed that challenges tackled have aligned themselves with these dimensions. This breadth has prompted us to emphasize the fallacy of the universal power model: no single power model can cover all seven dimensions through inclusion of variables and parameters. It is essential that prospective users of any such power model be aware of the limits of its scope. On the other hand, I have pointed out that the state-of-the-art includes adaptable modelling systems that handle variability in more than one dimension. Moreover, at least limited variability in two of the seven dimensions – workload type and concurrent operation of (multiple) VEs – is commonly validated, i.e., whether the model is truly capable of predicting power consumption under variability in workload type and number of concurrent VEs.

PAD elicits trends in its proceedings through a sample of a corpus. In particular, the following examples are among the most noteworthy (but not the only) saliencies.

1. The challenge category tackled most frequently is that of how to load the VC's resources and how to quantify and measure the load; disaggregating the predictors and attributing measured power consumption to the individual VCs is the second most frequently tackled.
2. The variety of approaches that tackle a (category of) challenge is positively correlated to the frequency with which it is tackled.

3. Instrumentation of computing resources (e.g., instrumentation of microarchitectural artifacts) is the most commonly adopted approach (towards developments), surpassing instrumentation of the workload.
4. Resource-specific workloads (e.g., processor-specific) are the most commonly utilized, surpassing workloads representative of real use (e.g., web applications).
5. In developments, the most commonly-developed model type is that where the power consumption (of the host or VC) is predicted in terms of workload characteristics; power consumption as a linear function of computing resources is second.
6. At the other end of the frequency range of developments, machine-learnt models comprise the second least frequent category of developed models, and adaptable, non-linear VC models are also very infrequent.

The process of parsing works and aggregating their codes is, however, only the principal ingredient in the overall progression towards the end goal: a set of themes that suitably profile the works in an area of research. Indeed, these codes and their inter-relationships have elicited several research gaps, pitfalls, and a fallacy, as well as evidence of the state of the art and of researchns.

7.5.2 A framework for development of real-time, predictive power models

Evolution of the research space on power consumption in virtualized environments now suggests the following framework for further development of power models:

1. Division of the problem into:
    a. a modeling concern:
        i. what components to include;
        ii. what workload(s) to consider;
        iii. what state factors (temperature, frequency, performance and idle states) to account for;
    b. an attribution concern, i.e., how to attribute host power to VEs
2. Division of the approach into:
    a. microarchitectural instrumentation, based on intimate knowledge of the microarchitecture and the memory system;
    b. granular attribution based on time-division multiplexing;
    c. model selection.
3. Development of parameterized models, subject to continuing (if not continuous) optimization of the parameters under machine learning.

Chapter 8. An implementational model spanning access node to metro-core: trends in, and motivations behind Communication Service Providers' activities

The outstanding part of the pursuit of this work's objectives may be pithily stated as concerning the development of an implementational model of the complement to the access network in the metro area, extending from:

- the access node[30] at one end, to
- high-volume packet switches interfacing to long-haul links at the other end.

This "complement" may be succinctly referred to as the AN – metro-core span. The overarching goal of the remaining work, is, therefore, ***to provide a representative sample of implementational models of the AN – metro-core span***, thereby:

- supplying energy analysts with directly usable implementational models, and
- enabling energy analysts to develop their own models for scenarios left out of the sample.

In chapter 5, the transmission media layer network was employed as the sole layer over which to superpose organizing reference points. It is necessary to complement this layer by use of other G.800 artefacts, in order to supplement the description of some of the representative architectures with ***client layers***. These client layers include components that consume energy and therefore should be drawn to the attention of the energy analyst. G.800 was first introduced in chapter 5; more extensive use of its modelling artefacts is required. This technique (i.e., use of G.800 modelling artefacts) is an important ingredient of the method, as it supports ***the need to draw attention to layers that are subsumed within the physical layer (layer 1) of the OSI seven-layer framework***. The data plane (or transport plane) is, due to recursion, a stack of layers: failure to acknowledge this recursion leads to misrepresentation of system boundaries. If the analyst does not wish to descend into all the nested layers, use of G.800 concepts will, at least, support accurate establishment of system boundaries. In essence: this technique in the method is "not fixed to a certain number of layers" [30].

While the method developed thus far has served well in laying out the access network, it is necessary to resume by stepping back to survey the problem domain from a greater distance. This detachment engenders a vision for a framework within which development of this span's implementational model can proceed. The product of these proceedings is presented in the next sub-section: in concrete terms, it consists of supplementation of the extant method by the techniques of

---

[30] The access node is a key demarcation between the access portion and its complement in the metro area network. It will be dealt with in detail in this chapter.

quantitative survey, qualitative survey and case studies within a well-rounded framework that integrates these techniques with those identified in preceding chapters. The supplementary scope – quantitative and qualitative survey – are dealt with in this chapter.

8.1 A three-axis framework for further development

The implementational model weaves ***topological components*** (introduced in chapter 5) and ***network functions*** (introduced in chapter 2) into an architecture suited to the purposes of the energy analyst. Development of the implementational model must investigate the evolution of these two principal aspects to accurately represent functional deployment within the metro area network.

The functions ***classically*** deployed in this part of the metro area are the following.

- Transport aggregation (while going upstream; it is distribution in the downstream): Aggregation is a well-distributed function of specific network elements en route upstream from the access node. Within the scope of the metro area network, traffic thus aggregated terminates at one end on the user equipment (UE) and at the other end on a network gateway, e.g., an Internet BNG, a video BNG, an EPC SGW (serving gateway), and, in 5G, a user-plane function (UPF).
- Service authentication, authorization and accounting (AAA): this is localized but the equipment, perhaps in the form of a broadband network gateway (BNG) may be deployed in several places. The BNG may implement the AAA functions of a video service private to the CSP, or it might implement the AAA functions for the global Internet.
- Traffic classification (for service differentiation): this has traditionally been delegated to access node aggregation (L2) switches, with the support further upstream of provider edge (PE) packet switches (L3).

In chapter 5, it was seen that these traditional functions are augmented by those that support 5G and its use cases.

1. 5G introduces the DU and the CU. Traditionally part of the baseband unit (BBU), and therefore located deep in
2. the access network, the functions subsumed within them, despite forming part of the ***radio*** access network, can now be virtualized, disaggregated from the radio unit and moved into a local exchange (which houses the access node). Therefore, while still lying downstream of the V RP, these functions have now moved upstream of the U RP. With the BBU, the functions subsumed within 5G's CU and DU lie downstream of the U RP. The CU may even be deployed further upstream of the V RP, which would take it outside the access network.

3. UPFs lie at the N6 interface to multi-access edge compute (MEC) nodes. Since MEC nodes are points of deployment of services, then any site of an aggregation junction is a suitable candidate for deployment of a UPF. Therefore, although the UPF is considered a part of the 5G core, its physical site of deployment may lie anywhere within the metro area at sites of aggregation. Therefore, these junctions must be carefully annotated in implementational models.

Topological components are evolving under three major pressures.

a) ***The development of technologies that disrupt established topologies.*** A useful example was first referred to in chapter 5. In XR optics, a frequency band is divided into (digital sub-) carriers; this facilitates the division of a single optical channel of large data rate capacity into several sub-channels. A passive splitter/combiner distributes the sub-channels to end-points for which lower data rates suffice. Thereby, the ***cost*** of distribution and aggregation is significantly reduced [25] (transceivers at the point of distribution/aggregation – the hub – are reduced from the number of end-points to one) through this point(one transceiver at the hub)-to-multipoint(one transceiver per end-point) arrangement.
b) ***The adoption of new technologies by communities of network operators (CSPs)***. The litmus test of technology is its adoption in the field. For example, while ATM used to be an aggregation architecture, its role has been supplanted by Ethernet (see, for example, the BBF's report on migration from ATM to Ethernet-based migration [149]). Each technology impacts either topological components, or location of deployment of network functions, or scope of deployment thereof, or any combination of the three.
c) ***Convergence of wireline and wireless networks, and convergence within wireline.*** Rationalization of diverse services onto a common transport infrastructure is a major motivational force for CSPs; the BBF's TR-470 (Issue 2) lists seven specific motives for convergence (of wireless and wireline) [80, Sec. 1.1]. Evidently, convergence re-writes all models; moreover, it is a major area of activity within ETSI, where Industry Specification Group (ISG) Fifth Generation Fixed Network (F5G) has, as of May 2023, published a network architecture with an explicit business requirement of "convergence and consolidation" of the various fixed networks [316, Sec. 4.2].

Therefore, three major axes for development of the implementational model can be perceived.

1. The implementation of classical functions is a mature field and current practice must be investigated and documented.

2. Due consideration of the three pressures' resultant thrust needs to be taken. These form trends that may be detectable; any detected trend must be documented and analysed for its causes and effects.
3. The paradigmatic shifts of 5G and MEC introduce new participants (e.g., related to computing embedded in the MAN), new functions and new interconnection points. An observation that 5G and MEC should be added to the three pressures would be a good one. I have separated them from the set of three pressures because the latter predominantly influence topological components of the AN – metro core span, while 5G and MEC mostly introduce new functions to this span of the metro area.

A terse summary of the framework would thus be: depart from the status quo and selectively, through investigation of trends, include emerging functions and topologies.

8.2 Further development of method

The framework of the three major axes guides the formation of a method that can be brought to bear on the problem of an implementational model for the metro area of a telecommunications network. This section expands the method developed in chapter 4 and chapter 5, as follows.

1. Sub-section 8.2.1 summarizes the essential threads of the method developed thus far, recapitulating previous development to facilitate continuity.
2. Sub-section 8.2.2 develops the method further, through techniques that align with the three major axes.

8.2.1 Recapitulation of the method thus far

The rationale of the method is summarized first, followed by the artefacts that emerge from its application. As regards rationale: it has been sought to extend extant standards that model telecommunications networks, to support the perspective of the energy analyst. This overarching approach bifurcates into consideration of model ***artefacts*** and model ***scenarios***.

1. **Artefacts** created by such standards have been assessed on two bases.
    a. **Comprehensive capture of energy consumers.** Suitability decreases with abstraction; for example, the reference point for interconnection – service (RPI-S) is unsuitable as it abstracts underlying network layers.
    b. **Breadth and consensus on adoption**. The more the number of SDOs that use an artefact, and the greater the agreement on its interpretation, then the more suitable it is for further extension of its purpose to meet the needs of the energy analyst.
2. The **scenario** is defined in ITU-T Y.120 as "a combined graphical and textual representation of [a configuration of] … network technologies and user appliances that

may be expected to be encountered in the context of the Global Information Infrastructure" [183]. Scenarios are expected to support resolution of boundaries of service provision, facilitate system integration (since there are participants from diverse backgrounds) and assist in the identification of what needs to be standardized. **Scenarios** have been selected on the basis of the distinction between the transport functional group and the control functional group, since the transport functional group carries consumer traffic and is thus the dominant energy consumer.

The artefacts that emerged from this approach are the following.

1. The **implementational model** defines deployment of network functions.
2. **Reference points (RPs)** divide functional groups. Therefore, they are good candidates for the role of demarcation of a system's boundary.
3. **Reference points for interconnection – network (RPI-Ns)** define points of interconnection between **networks**, and therefore support identification of **physical** separation of functions.
4. **Layer networks** (ITU-T's Series G of recommendations[31]) support the recursion of transport systems down to the transmission media layer network. This progressively resolves abstractions until all energy consumers – in "transfer (moving bits), transform (computing) and storage" [317] – are accounted for. Layer networks are the means for vertical (up and down) traversal of nested networks. ITU-T G.800 [175, Ch. 6] states that "[i]t is recommended that this method[32] be used to describe the transport network".
5. **Inter-domain interfaces (IrDIs)** and **Intra-domain interfaces (IaDIs)** [184] are a special, and potentially highly significant, sub-class of RPI-N. They specify interfaces for optical networks; an optical network is a specific type of **transmission media layer network**. Therefore, IrDIs and IaDIs, while not essential, nor indeed even always possible (e.g., where RPI-Ns are established on non-optical media), guarantee full concretization of architectures down to equipment chassis boundaries. They provide a means for horizontal partitioning of the optical-network layer-network and, by extrapolation, potential RPs for horizontal partitioning of client layer networks.

---

[31] Series G: Transmission systems and media, digital systems and networks

[32] That is: use the layer network, not the OSI layer, to describe the transport network.

8.2.2 Major techniques that align with the major axes

1) CSP survey – aligned with first and second major axes

The first major axis regards an investigation of implementation of classical network functions. Furthermore, the second major axis includes an investigation of CSPs' adoption of new technologies. Both these components of the framework can be studied by surveying CSPs. A quantitative survey was prepared to gather statistics. However, questionnaires used in quantitative surveys require careful design to withstand criticism. The following divisions of this sub-sub-section first deal with the criticism of the method, then with defence against the criticism and practical implications for the method, before describing the survey itself.

a) Criticism

A reputable network engineer whom I contacted, but wishes to remain anonymous, opined that "[q]uestionnaires are collections of problems that are poorly defined, answered by someone who doesn't understand them, and even if they did, they don't really know what their company is doing and is going to be doing going forward" [318]. In point form, the objections are:

1. questionnaires are poorly defined;
2. respondents answer regardless of whether they understand the questions or not, and
3. respondents are unaware of the status quo in operations and their organizations plans for development of operations.

b) Defence

The implications of the criticism are severe for researchers at least partially dependent on this research tool; it was necessary to understand the **validity** of these objections. If valid, then it would prove necessary to **safeguard** against them.

On the validity of the objections

A highly experienced, well-known analyst (Sterling Perrin[33]) in the field of telecommunications, who regularly employs questionnaires, and with whom I am acquainted, was contacted to understand the general **validity** of these objections [319]. The analyst's response is cited first, then paraphrased.

> *"I think a well-developed survey mitigates most of what he stated. We spend a lot of time/effort in crafting the questions so that they are not ambiguous and are well-defined (e.g., including specific definitions). The database(s) is important because we solicit responses from people who are*

---

[33] https://www.linkedin.com/in/sterling-perrin-2492b5/

> *directly involved working in the survey topic and have an understanding of their company's plans in the topic area. He is describing a very poor survey. I've described the ideal survey – the reality is probably somewhere in between.*
>
> *We individually scrutinize each response to remove suspicious and uninformed responses and push the aggregate as close to the ideal scenario as possible. Again, not perfect, but in aggregate I have found survey results a reliable indicator of major market trends and shifts. The more granular we cut the data, the more likely to obtain results influenced by a bad response in the numbers. We aim for 80-100 responses for this reason, and rarely analyze a cut less than 40 (and if so, with a disclaimer that the size is small).*
>
> *A couple of other observations: 1. consumers don't really plan but businesses do. It's reasonable to ask business employees about their plans for their business, because they spend a lot of time planning for the future. Plans change, but a survey will capture plans at a moment in time. 2. Businesses (and consumers) understand their pain points. This data is very reliable. 3. Businesses also tend to understand the trends influencing their business, so surveys are good indicators here. 4. Surveys (in my experience) have been least accurate in identifying timing. They know what they want, but it takes longer than expected almost always. I rely on timing data least of all."*

This can be condensed into the following five recommendations for a researcher designing a quantitative survey of the population of CSPs.

1. Major market trends and shifts are reliably predicted by quantitative survey; narrower interests are less reliably predicted.
2. Write unambiguous questions, using definitions where necessary.
3. Ask people whom you know are qualified to respond; here, this instantiates to:
    a. ask people who are involved in network operations, and
    b. ask people who understand their organization's plans in network operations.
4. Individually scrutinize responses for evident inconsistencies, and eliminate such responses from the set of responses.
5. A sample of 80 – 100 responses has been found to provide statistics that match the population's (CSPs) parameters well.

Since the validity of the objections was established, it proved necessary to safeguard against them.

On safeguarding

Safeguarding was tackled as follows:

- **Pre-survey**
    - through support in writing the questions, before dissemination of the questionnaire:
        - the questions were reviewed by Seacom's head of engineering (Mark Tinka[34]) [320], and
        - demographic questions were suggested by an analyst experienced in writing questionnaires [321];
    - through use of names of standards that uniquely identify technology, where necessary (e.g., rather than write "GPON" to refer to this access technology, "GPON (ITU-T G.984.1)" was used);
    - through use of graphics from standards documents where specific points in the metro area were invoked;
    - technical questions did not mandate an answer; only questions that addressed demographics were mandatory. Indeed, as the results show, most of the answer sets (one set per question) carry less responses than the number of respondents.
- **Post-survey**
    - through scrutiny of individual responses [322];
    - through review using the means of a qualitative survey[35] (see part (d)).

c) The quantitative survey

The questions

The questionnaire underlying the quantitative survey has six optional, technical sections concerning metro area network architecture and one mandatory section concerning demographics. The organization of the technical sections is illustrated in Fig. 64. A brief description of the sections follows.

1. The first two (technical) sections concern size and rate of growth of access technologies by subscription, while the third concerns architecture of access technology implementation.

---

[34] https://www.linkedin.com/in/mark-tinka-5b03055/

[35] The scope of the qualitative survey extends beyond assessment of the objective clarity of the questions.

2. The fourth concerns the network services that CSPs are selecting to deploy the various stages of RAN traffic hauling (fronthaul, midhaul and backhaul), and the dissemination of use of disaggregated cell site gateways (DCSGs).
3. Aggregation (beyond the V RP) is dealt with in the fifth technical section. Concern here lies with trends and shifts in the optical network (layer 0), the physical layer (layer 1) and the link layer (layer 2).
4. The sixth section is aimed at an investigation of the adoption of deep service edges (i.e., service edge closer to the subscriber than the local exchange/central office/HFC hub).

A sample of the questionnaire may be found online[36].

The sample

Two samples were collected. One of the samples was obtained through market research, conducted by SG Analytics. SG Analytics has conducted several studies with CSPs. SGA's pool of respondents comprises all types of CSPs across the globe. The CSPs are differentiated based on the services they offer and the size of their company. For this study, global and regional CSPs from the pool like AT&T, T-Mobile and Verizon were targeted. Respondents were vetted by SGA before being included in the sample. SGA populates its pool of individuals from sources such as LinkedIn, Zoom info, conferences, and magazines. Data on pool individuals is processed using general criteria; individuals who pass these then proceed to project-specific criteria.

---

[36] https://forms.gle/QtoTkhzEk4Q1BLdVA

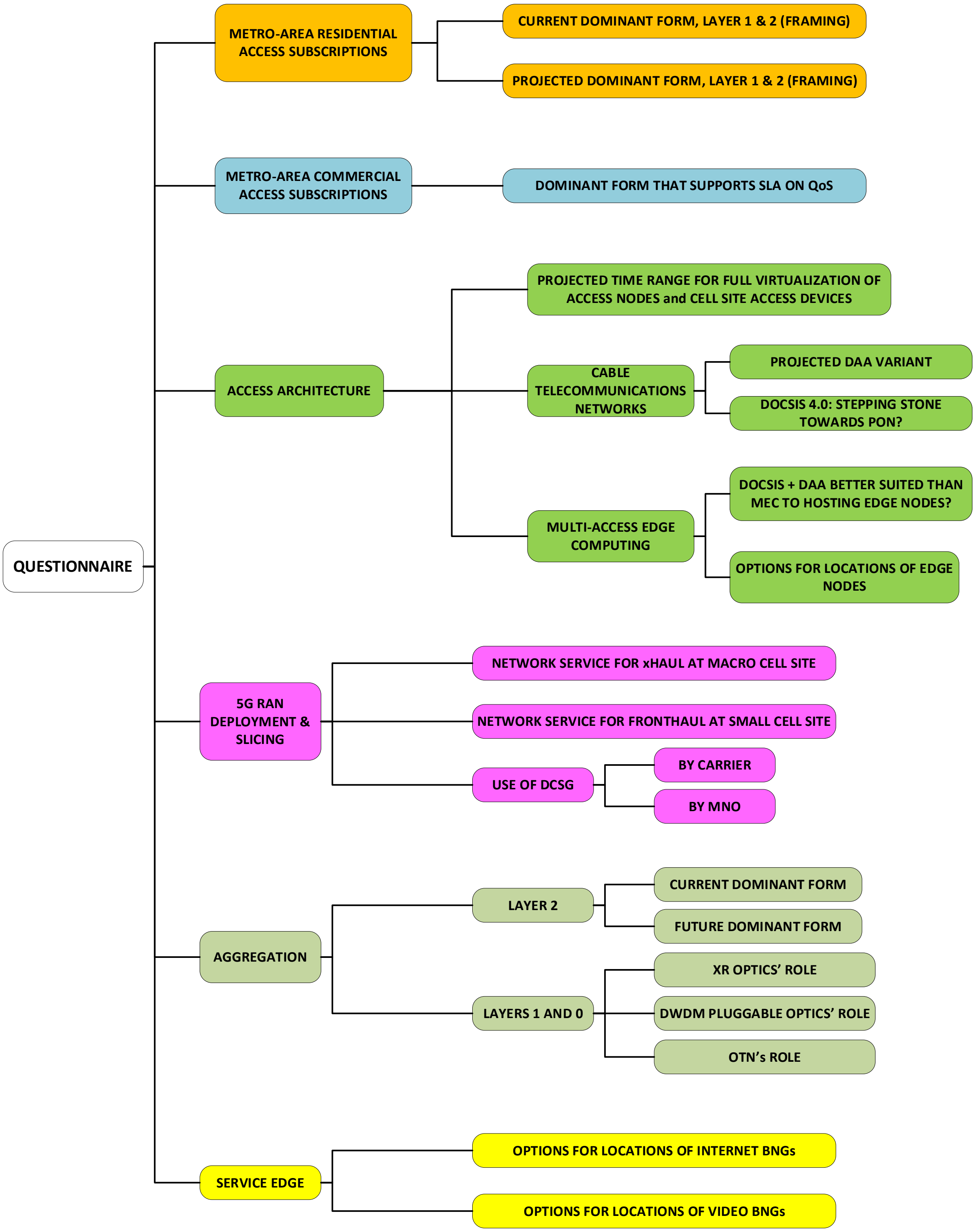


Fig. 64. Structure of the questionnaire used in the quantitative survey

- **General criteria**
  - ***Demographics*** Pool information is updated every three months, to capture the most recent status of the pool individual member's demographic details. The primary focus is on the individual's decision making in his/her current organization.

    - ***Behavioural patterns*** Every pool individual member is clubbed by SGA's scoring system, in order that his/her behavioural patterns on past projects may be assessed. Behavioural patterns are monitored by SGA's panel team on a regular basis.
    - ***Anomalies in responses*** SGA runs validation checks periodically on a pool individual member's historical data, to find anomalies in responses. Any pool individual member who fails these checks is removed from the pool.
- **Project specific criteria**
  For a specific project (such as this one) a pool individual member passes from pool to sample through a double opt-in, thereby becoming a **respondent**. Respondents pass through:
    - initial screening parameters that are set on SGA's panel for the project, and
    - my screening criteria (demographics section, excluding subscriber base size).

The general and the specific criteria are designed to select the best-fitting pool members to participate in the survey. Furthermore, individual responses were read and those which presented inconsistencies (e.g., fastest-growing access technology was not among the set of deployed access technologies) were removed from the sample.

The second sample was collected from operator groups'[37] mailing lists around the world. Respondents were not authenticated, but only one fake response was found and considerable effort was required to answer the entire set of questions in a coherent and cogent manner. In this latter regard: during qualitative review, Philip Smith observed that "we are stuck with those who are willing to volunteer their time" [323]; the time estimated by SGA's personnel for a response to the complete set of questions is between 10 and 12 minutes [324].

The quality of the questionnaire

The five recommendations are now brought to bear to assess the quality of the questionnaire.

1. **Major trends or minor movements?** The finest detail in the survey is found in the section asking about aggregation in the optical network. One question there asks about motives for migrating towards optical networks that use DWDM transceivers that are directly pluggable into router chassis faceplates. Otherwise, all questions regard differentiating between major technologies. For example, the aggregation section asks about current and future adoption of optical transport network (OTN, [325]) and about

---

[37] In alphabetical order: AFNOG, APOPS, AUSNOG, DENOG, ENOG, FRNOG, GORE, IDNOG, INNOG, ITNOG, JANOG, LACNOG, NANOG, SAFNOG, SANOG, SWINOG, UKNOF

whether provider bridging (Q-in-Q) is preferred over seamless MPLS transport; the service edge section asks about current locations where Internet BNGs are in use (and illustrates the locations using a graphic from the BBF's TR-178).

2. **Write unambiguous questions, using definitions where necessary.** Apart from support during crafting, the qualitative survey (described later) asked interviewees specifically about difficulties of this kind. Three issues emerged.
    a) The term "routed optical network" is proprietary to Cisco. A discussion was opened in this regard on NANOG [326]. Eduard Vasylenko[38] (Huawei) correctly identified this as term as proprietary ("[n]obody understands what you mean"); Ovidiu-Mădălin Roşeț, a CCIE (Cisco Certified Internetwork Expert) opined that "I don't think they thought about Cisco's routed optical networks when they chose this" [327, N. @35:23]. These observations are mitigated by one from an interview [328, N. @8:36] (matching my personal experience) with another CCIE (Haider Khalid) in the qualitative review phase, that "Cisco is actually the go to vendor for everybody". The consequences of this ambiguity are handled in the interpretation of the results.
    b) "Segmented (as opposed to seamless) MPLS transport" might not be universally understood; a doubt emerged in [327, N. @4:41]). However, the segmented pseudowire is well-defined [329], and use of the term in contraposition to seamless MPLS should have sufficed to distinguish the meaning intended. Moreover, in "draft-ietf-mpls-seamless-mpls-07" [330, Sec. 2.1], a clear contraposition is made between the status quo ante and the proposal for seamless MPLS.
    c) XR optics (see reference made earlier) is sufficiently novel to evade correct identification. During a review with Jon Baldry[39], Infinera's "Director Metro Marketing", doubt was cast on the claim that two respondents from the NOG sample had already deployed the technology, as he was unaware of any commercial deployments. A clarifying reference was added to the question, but data collection was by then roughly at two-thirds the final sample's size.

From a more general perspective: two reviewers were explicitly invited to comment about limitations and ambiguities in the questionnaire's questions [331], [332]. They explicitly stated that they did not see any ambiguities. Both suggested

---

[38] https://www.linkedin.com/in/eduard-vasylenko-b723ab1

[39] https://www.linkedin.com/in/jon-baldry-5605b/

some extension of scope of the questionnaire, but these were concerned with the control plane rather than the data plane.

3. **Ask qualified respondents**. The problem of qualifying respondents was dealt with through two approaches.
    a) Relevant organizational roles were specified (see the demographics section of the questionnaire in Appendix 5[40]).
    b) A vetted database of respondents was used in the first sample.
4. **Scrutinize individual responses and eliminate those with inconsistencies**. Every response in both samples was subjected to scrutiny. This process is documented for the first sample in [322].
5. **Sample size of 80 – 100 responses**. The first sample was restricted to 50 responses because of budgetary constraints, with 30 different companies represented. The second sample consists of 79 responses; this should still suffice.

d) The qualitative survey

The objectives may be broadly classified under two headings.

1. Discuss the graphical summaries of the results of the quantitative survey.
2. Assess the objective clarity of the questions in the questionnaire.

Qualitative survey was carried out using the means of:

1. Face-to-face interviews, which were recorded;
2. e-mails, and
3. a written assessment.

Face-to face interviews

1. Both the objective classes were tackled during the interviews.
2. Three participants were recruited.
    a. Two were recruited by SGA through a process that included filtering by a set of screening questions which I wrote (see Appendix 6) with the support of the anonymous network engineer.
    b. The latter's suggestion (in [318]) to recruit knowledgeable people with direct participation in the field of network operations – colloquially, we would say that their hands are dirty – was supported by Mark Tinka. My own experience gives

---

[40] https://forms.gle/QtoTkhzEk4Q1BLdVA

credence to this opinion; to strengthen it further, I qualified it (in the screening questions) by the criterion that prospective interviewees should be experienced.

3. All interviews were recorded [327], [328], [333].

E-mails

1. Both the objective classes were tackled in the course of the exchanges in the mail threads.
2. Three highly-experienced and highly-qualified participants shared their interpretations:
    a. Mark Tinka[41];
    b. Philip Smith[42];
    c. Daniel King[43];
    d. the anonymous reviewer referred to earlier.
3. All interactions are documented and available [318], [323], [332], [334].

Written assessment

While interviews are a useful medium, they are limited by the time which all parties involved can endure in discussion and remain at sufficient ease to interpret graphical summaries of data. One of the interviewees accepted the request to follow up with a written assessment.

1. Both the objective classes were tackled during the written assessment.
2. Haider Khalid, a highly-qualified network engineer participated.
3. The assessment is documented and available [331].

2) SDO activities and Case studies

The CSP survey explores the first major axis and the thrust expressing CSPs' adoption of new topological components (2nd axis). It now remains to explore the thrust exercised by disruptive aggregation technologies, the thrust exercised by wireless-wireline convergence topological components and the interconnection point of new 5G functions and new MEC functions. Development of method to address these axes includes:

1. **Case studies** (2nd axis). These shed light on the significance of disruptive aggregation technologies.
2. **SDO activities.** Activities in scope are those that relate to foundational architectures.
    a) Wireless-wireline convergence (2nd axis): Reference has already been made to ETSI ISG F5G's work, which is still under development. The BBF has published its second issue of TR-470, "5G Wireless Wireline Convergence Architecture"

---

[41] https://www.linkedin.com/in/mark-tinka-5b03055/
[42] https://www.linkedin.com/in/philip-smith-154502/
[43] https://www.linkedin.com/in/danielking/

[80]. Jointly, these works help explore the thrust exercised by wireless-wireline convergence.

b) 5G functions (3rd axis): the ITU-T's G.8300 [174] refers to alternatives for the location of the 5G functions within transport in the metro area.

c) Multi-access edge computing functions (3rd axis):

  i. ETSI's MEC reference architecture [335] standardizes functions and reference points, and

  ii. BBF's use cases report for metro compute networking [336] supports location of MEC functions within transport in the metro area.

8.3 Quantitative survey results

This section presents a graphical summary for each question that concerns the AN – metro-core span, and for each question that concerns the service edge and 5G functions. Since MEC nodes (these form part of the deployment range for the service edge) and 5G functions may be deployed within the access network, questions that regard them (the MEC nodes and 5G functions) may pertain to the access network. The graphical summaries consist of histograms, bar charts, clustered bar charts and pie charts. Brief commentary accompanies the charts, to draw attention to noteworthy characteristics. Analysis is deferred to section 8.5, where, aided by reviewers' comments, decisions on which scenarios to model are taken (sub-section 8.5.1).

8.3.1 The samples

This sub-section first presents essential statistics on both samples. It then proceeds to extract a correlation from the adoption of technologies in the access network (this segment is both CapEx and OpEx intensive); this correlation is important because it suggests an interpretive bias for the rest of the results. The sub-section concludes with a summary of the results, in Table XVII .

Essential statistics

Table XVI ESSENTIAL STATISTICS ON SUBSCRIBER BASE

	NOG sample (thousands of subs.)	**SGA sample (thousands of subs.)**
Mean	5510	42887
Median	100	30687
Mode	1 (28 instances)	101 (2 instances)

A correlation between subscriber base size and access network technology

For brevity's sake, the sample collected from the operator groups will be referred to as the NOG sample, while the sample collected from SGA's database will be referred to as the SGA sample. Figures will be presented with the NOG chart on the left and the SGA chart on the right. The results of the demographics questions are useful to understand respondents' organizations better. Fig. 65 carries two histograms of size of subscriber base. Size of subscriber base was obtained through a conservative estimate. For example, consider Fig. 66. Suppose that a respondent selects the radio button corresponding to a choice, say, in the range one thousand – one hundred thousand subscribers in one – ten metro areas. Before aggregation with the other answers for the other geographical areas, this was converted (for the region of North America) to one thousand subscribers in one metro area.

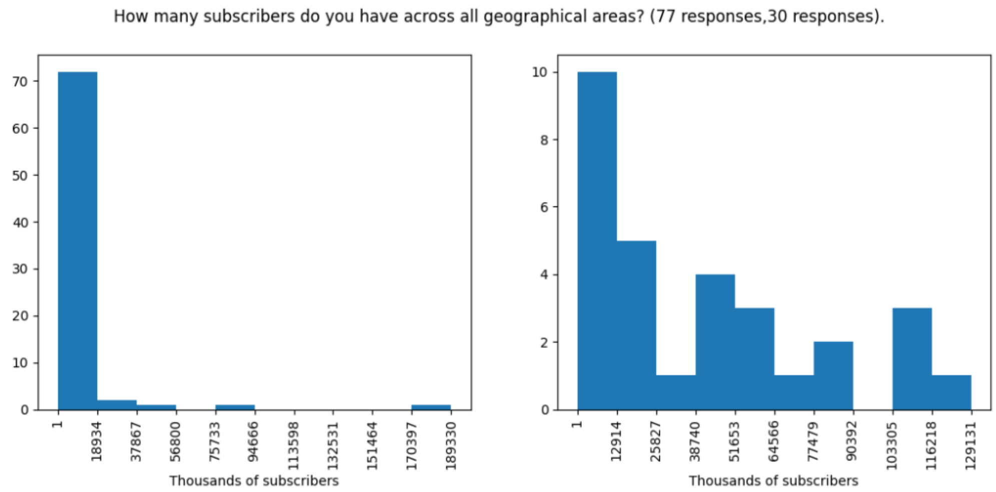
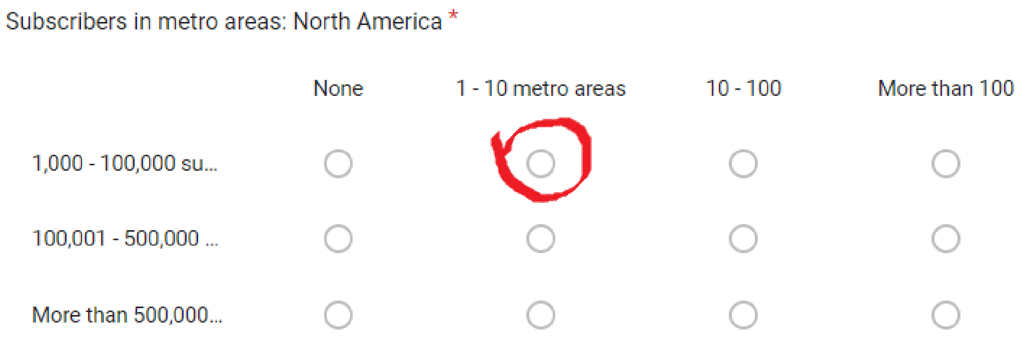

How many subscribers do you have across all geographical areas? (77 responses,30 responses).

Fig. 65. Distribution of a conservative estimate of size of subscriber base


Subscribers in metro areas: North America *

	None	1 - 10 metro areas	10 - 100	More than 100
1,000 - 100,000 su...	○	[✓ ○]	○	○
100,001 - 500,000 ...	○	○	○	○
More than 500,000...	○	○	○	○

Fig. 66. Choice of 1,000 – 100,000 subscribers in 1 – 10 metro areas is reduced to a conservative value of 1,000 subscribers

This conservative approach proved very useful in explanation of differences between the responses from the two samples. The sizes reported by NOG respondents are heavily skewed towards the low end of the range [1000, 189330000]. The sizes reported by SGA respondents are less heavily skewed towards the low end of the range [1000, 129131000]. This indicates that while SGA respondents do indeed (as claimed by SGA) include regional operators, the NOG respondents have smaller subscriber base size and more likely to be later entrants in the population of CSPs. This observation is reinforced by the questions regarding access technologies; while these are clearly not within the scope of the AN – metro-core span, the access network is the largest sink of capital and operational expenditure. Fig. 67 shows the adoption of access technologies by respondent. Inspection of Fig. 67 reveals that the NOG respondents – the smaller CSPs – claim that GPON is both the most adopted access technology (blue bars), as well as the one that is growing at the fastest rate (orange bars). On the other hand, the larger operators – captured in SGA's sample – identify ADSL2+ as most adopted and fastest growing access technology.

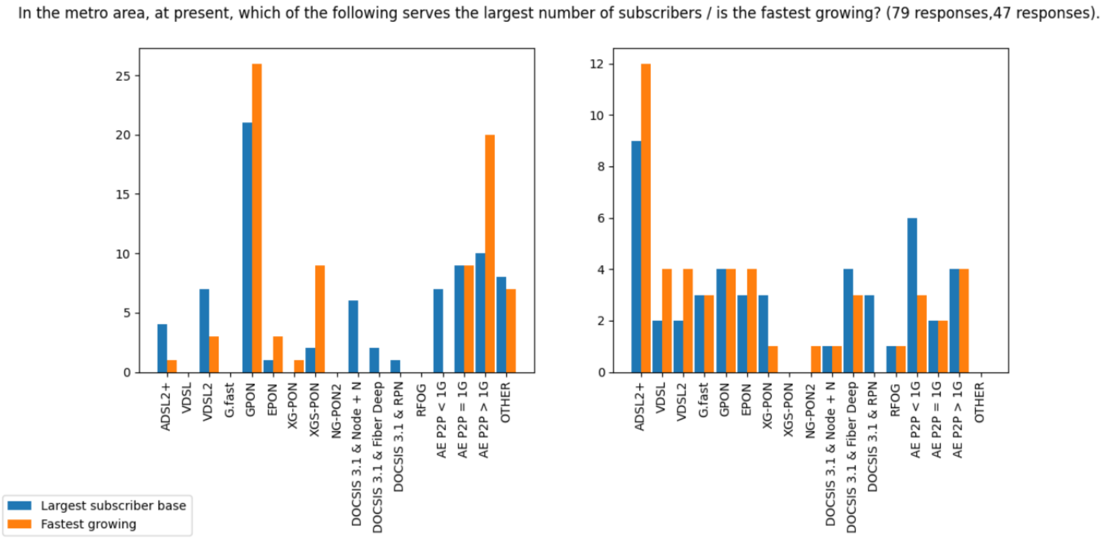


Fig. 67. Access technology adoption, unweighted

This result merits a deeper look before the full analysis is dealt with, as it serves as part of the interpretive lens through which the survey's data is to be read. Consider Fig. 68[44], which shows a ***weighted*** version of the responses, i.e., each response is multiplied by a factor numerically equal to the conservative estimate of the size of the subscriber base. Clearly, not all a respondent's subscribers are on the same access technology; at the same time, if a respondent indicates that an access technology is the respondent's most adopted access technology, then further light might be shed.

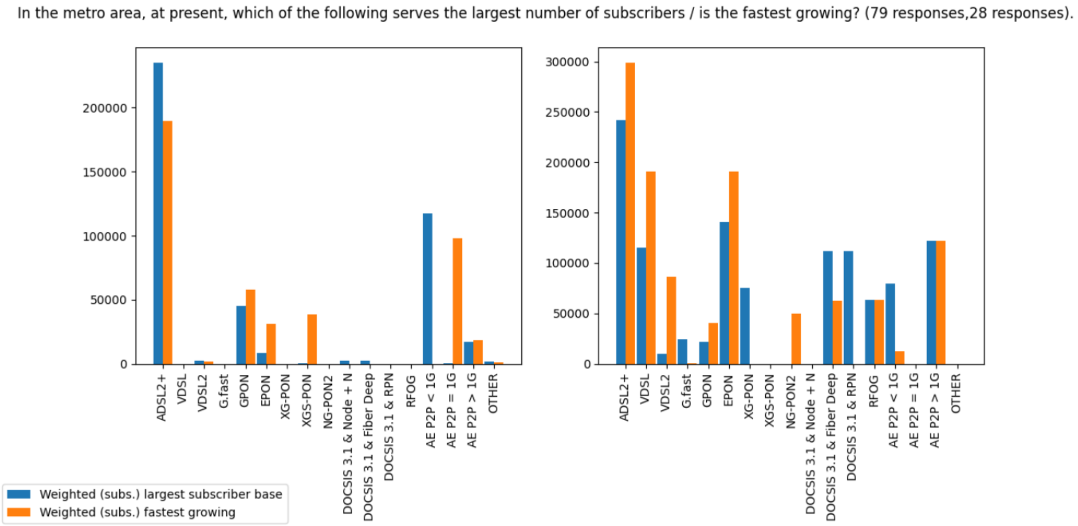


Fig. 68. Access technology adoption, weighted by size of subscriber base

---

[44] The figure only shows 28 respondents in SGA's sample; these correspond to the 28 different CSPs in the sample.

In the weighted set, ADSL 2+ is the largest and fastest growing in both samples. This indicates that the more established CSPs (including incumbents) are still largely exploiting their investments in copper media. This conclusion was reached following discussion with SGA and an expert reviewer [334]. Therefore, further interpretation of the data can be supported by the understanding that respondents in the NOG sample are likely to be less bound by legacy than those in the SGA sample. This understanding was affirmed in discussion with another expert reviewer [323].

> *"Doing the survey by open request to the NOG community will mean you'll get the smaller operators and the operators who are not driven by their vendors – the big private operators for example. Is it representative? Well, who knows, but we are stuck with those who are willing to volunteer their time. Again, I've helped these providers over the years, and they are much more determined to do what is right for the customer, the best and most reliable implementation, most cost effective for them to implement and operate. (I've been advising one just now on this very thing – they want particular tech for a new mobile access network, but unfortunately their management signed up for a vendor first – the tech team now has no say in what is being deployed because the vendor has a "solution" they are going to deliver. I'm looking at the technology choices and shaking my head – it's going to harm this operator's potential for success as their competitors are already looking at the newer technologies.)"*

A quick-and-dirty differentiator: Tier 1 vs "regional" vs "incumbent"

The role played by "Tier 1" CSPs is not identical to that played by the incumbents. Tier 1 CSPs provide transit to other CSPs and to service providers (e.g., video) who operate OTT (see Fig. 3), as well as to Cloud data centres. Incumbents, as stated earlier, are most likely to be CSPs who have invested heavily in structural and infrastructural works, possibly being the privatised descendant of erstwhile state-owned enterprises for national telecommunications. Incumbents may be restricted to regional (perhaps even national) operation. On the other hand, the Tier 1 CSPs, while quite possibly descended from an operator of regional/national scope, have a broader geographical (possibly global) scope. Examples of regional/national operators who have evolved to include the Tier 1 role, are Telefónica and Deutsche Telekom, but there are newer entrants like Colt who do not share the same origin story. The latter do not have a significant residential service (this emerged from a direct interview with Colt's senior technical personnel), whereas the former two do. Moreover, subscriber base size is not equivalent to revenue size. While residential subscriptions are well-known to be low-margin, the same cannot be said about enterprise services.

Tabulation of results of quantitative survey of CSPs

Table XVII SUMMARY OF RESULTS

ACCESS ARCHITECTURE

Access node virtualization

GROUP	*Done*	*<=1 y*	*<=5 y*	*No plans to fully virtualize*	*Other*	*Responses*
NOG	11	9	11	65	4	75
SGA	30	49	21	0	0	43

DAA option

GROUP	*R-OLT*	*R-OLT conditional*	*RMN*	*RMN conditional*	*RPN*	*RPN conditional*	*Undecided, but within 5 years*	*No DAA*	*Other*	*Responses*
NOG	10	17	1	7	10	6	15	44	4	72
SGA	21	38	38	23	17	23	13	2	0	47

DAA for majority HHP?

GROUP	*<=2 y*	*<=5 y*	*Greenfield only*	*No*	*Other*	*Responses*
NOG	7	13	12	65	3	68
SGA	53	36	9	2	0	47

Option 0 MEC node

GROUP	*In progress*	*<=1 y*	*<=5 y*	*No plans to deploy*	*Other*	*Responses*
NOG	16	11	11	61	2	57
SGA	24	56	20	0	0	50

Services for fronthaul at macro cell sites

GROUP	*MEF service*	*MPLS service*	*PON ONU*	*Wavelength*	*Dark fibre*	*Wireless*	*Other*	*Responses*
NOG	14	32	5	12	32	2	4	56
SGA	24	35	16	8	3	14	0	37

Services for midhaul at macro cell sites

GROUP	*MEF service*	*MPLS service*	*PON ONU*	*Wavelength*	*Dark fibre*	*Wireless*	*Other*	*Responses*
NOG	11	39	2	13	27	2	7	56
SGA	22	24	5	19	8	22	0	37

Services for fronthaul at small cell sites/fixed wireless access

GROUP	*MEF service*	*MPLS service*	*PON ONU*	*Wavelength*	*Dark fibre*	*Other*	*Responses*
NOG	15	36	7	12	27	3	59
SGA	49	32	8	5	5	0	37

AGGREGATION ARCHITECTURE

Current dominant form of L2 or greater aggregation from access node (V RP) to service edge

GROUP	*PB, w/o MPLS*	*Seamless MPLS*	*Segmented (not seamless MPLS)*	*PB near access, then MPLS*	*Other*	*Responses*
NOG	19	43	11	18	9	74
SGA	49	28	15	9	0	47

Preferred form of L2 or greater aggregation from access node (V RP) to service edge

GROUP	*PB, w/o MPLS*	*Seamless MPLS*	*Segmented (not seamless MPLS)*	*PB near access, then MPLS*	*Other*	*Responses*
NOG	14	51	8	17	10	71
SGA	43	30	19	9	0	47

Do you plan to deploy XR optics in your metro-aggregation network?

GROUP	*Already deployed*	*By end 2022*	*By end 2023*	*By end 2025*	*Currently investigating*	*No plans*	*Responses*
NOG	4	1	7	5	27	55	74
SGA	0	26	21	6	40	6	47

Existing OTN aggregation wil stay in my network but I won't choose OTN for any expansion of my aggregation network.

GROUP	*Fully disagree*	*Somewhat disagree*	*Somewhat agree*	*Fully agree*	*Other*	*Responses*
NOG	9	33	46	12	0	57
SGA	4	2	70	23	0	47

If you agree that you won't include OTN, why not?

GROUP	*Cost*	*Granularity of b/w*	*Inability to meet 5G URLLC*	*Other*	*Responses*
NOG	52	19	21	8	36
SGA	48	29	22	0	44

Packet-based networks will fully displace OTN from MANs, except in DCI.

GROUP	*Fully disagree*	*Somewhat disagree*	*Somewhat agree*	*Fully agree*	*Other*	*Responses*
NOG	13	23	52	12	0	60

SGA	0	9	72	17	2		47

Current dominant form of technology stack in metro aggregation

GROUP	DWDM+SDH+E+ IP/MPLS	DWDM+ROADM+ OTN+E+IP/MPLS	DWDM+ROADM+ E+IP/MPLS	DWDM+ROADM+ E+IP	Routed optical nets. over E without ROADM	Other	Responses
NOG	10	13	14	6	45	12	69
SGA	23	23	28	13	13	0	47

Greenfield form of technology stack in metro aggregation

GROUP		DWDM+ROADM+ OTN+E+IP/MPLS	DWDM+ROADM+ E+IP/MPLS	DWDM+ROADM+ E+IP	Routed optical nets. over E without ROADM	Other	Responses
NOG		12	33	9	43	3	67
SGA		36	26	26	13	0	47

Greenfield form of technology stack in metro core

GROUP		DWDM+ROADM+ OTN+E+IP		DWDM+ROADM+ E+IP	Routed optical nets. over E without ROADM		Responses
NOG		10		37	49	3	67
SGA		36		40	23	0	47

SERVICE EDGE

Service edge location for Internet BNG

GROUP	Option 0	Option 1	Option 2	Option 3	At A10	Other	Responses
NOG	19	33	46	27	19	0	32
SGA	4	36	44	34	8	0	50

Service edge location for Video BNG

GROUP	Option 0	Option 1	Option 2	Option 3	At A10	Other	Responses
NOG	3	31	41	25	16	0	32
SGA	10	38	36	30	10	0	50

Support for eMBB is improved by adding video BNGs closer to the end user

GROUP	Fully disagree	Somewhat disagree	Somewhat agree	Fully agree	Other		Responses
NOG	2	22	50	20	5		40
SGA	0	4	48	48	0		50

Support for eMBB is improved by adding video BNGs closer to the end user

GROUP	*Fully disagree*	*Somewhat disagree*	*Somewhat agree*	*Fully agree*	*Other*	*Responses*
NOG	13	21	47	15	4	47
SGA	0	4	56	40	0	50

Is Carrier Ethernet most adopted service on UNIs subject to QoS SLA?

GROUP	*Yes*	*No*	*Responses*
NOG	67	33	78
SGA	41	2	43

8.3.2 Access Architecture

What is the time range within which you plan to virtualize all your access nodes (vOLT vs OLT, vCMTS vs CMTS) and/or cell-site access devices (disaggregated cell site gateway (DCSG) vs cell-site router (CSR))?

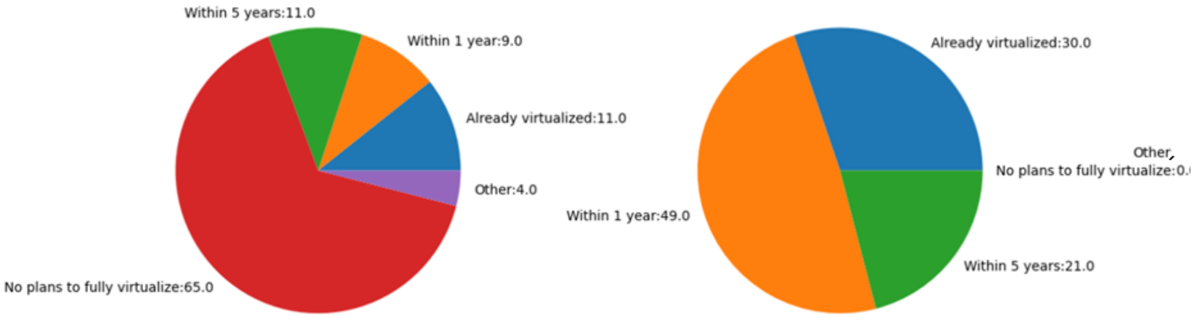


Fig. 69. Time span within which all access nodes and cell-site access devices will be fully virtualized (numbers show percentage of sample size)

Most of the NOG respondents do not plan to fully virtualize, while all of the SGA respondents plan to fully virtualize. This may reflect the relationship between the size of a CSP's employee cohort, and the cohort's skill set specialization. Larger CSPs are likely to employ more people and are likely to support greater specialization.

Which distributed access architecture (DAA) option(s) are you planning for new deployments and replacement deployments?

Fig. 70 shows a bar chart of the NOG sample on top and one for the SGA sample below it. A pattern emerges similar to that which emerges from Fig. 69. Smaller operators are far less likely to commit to DAA than the incumbents. No other characteristics is readily perceived from the data. Technology choices are fairly equally distributed among all the candidate architectures.

Do you plan DAA to serve the majority of your households passed (HHPs, when compared with centralized access forms such as centralized OLT and integrated CCAP)?

Fig. 71 concurs with Fig. 70, in the sense that while 43% of the NOG sample do not plan to deploy any DAA option, even more (65%) do not plan to make it the majority access architecture. On the other hand, the overwhelming majority of incumbents (89%) plan to do so within 5 years.

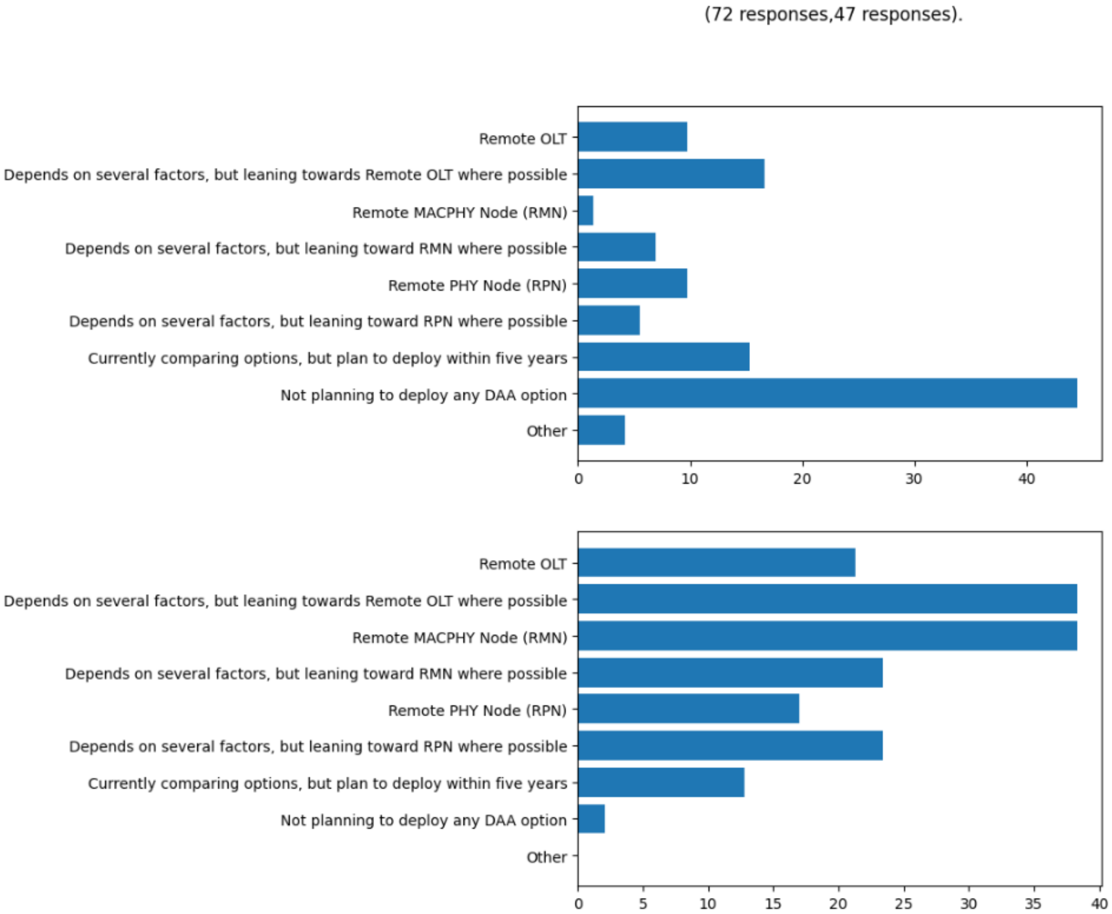


Fig. 70. Adoption of distributed access architecture technology (numbers show percentage of sample size, NOG on top)

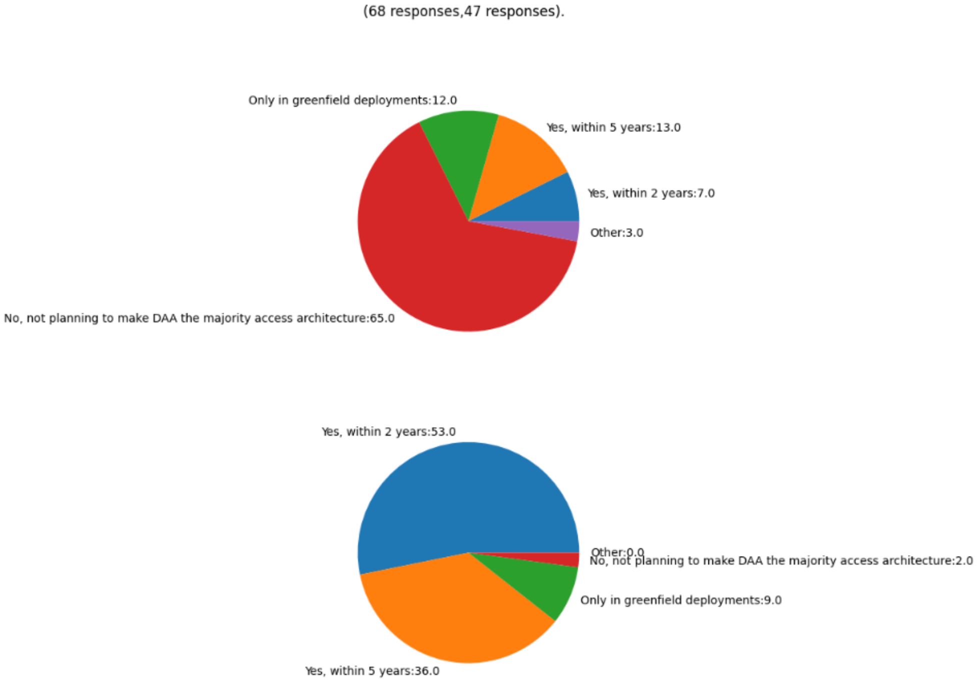


Fig. 71. Will DAA be your majority access architecture in HHPs? (numbers show percentage of sample size, NOG on top)

Do you plan to deploy remote access nodes (Option 0) to enable MEC services?

Option 0 (Fig. 72, [337, Fig. 2]) supports the greatest 5G and MEC functional range, but widespread provision is very capital-intensive (due to numbers). Results in the NOG sample (Fig. 73) are consistent with those shown in Fig. 71; similarly, SGA respondents seem intent on exploiting DAA real estate to deploy customer-proximal computing.

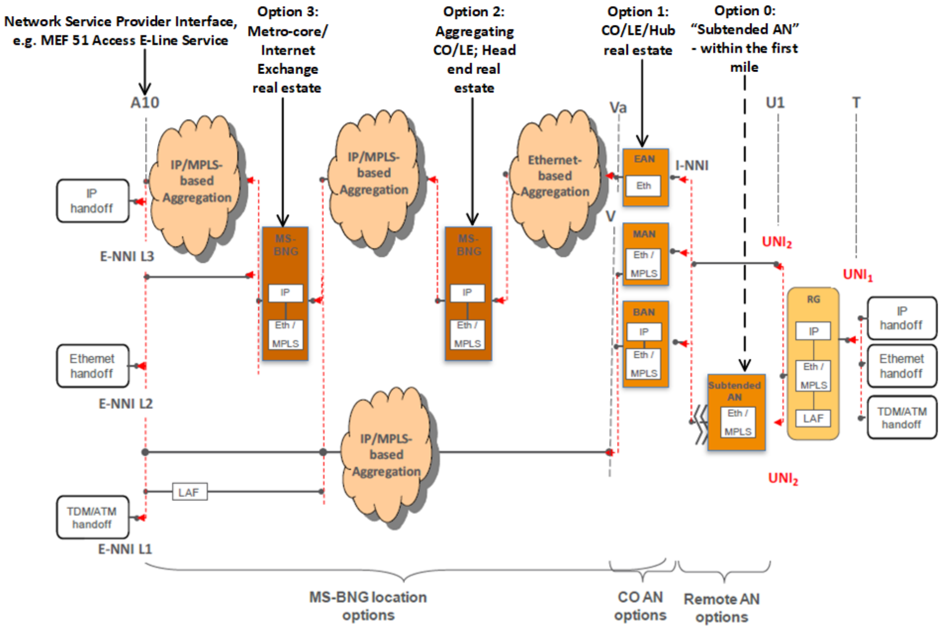


Fig. 72. General TR-178 architectural scheme, encompassing its targeted deployment scenarios [337, Fig. 2]

(57 responses,50 responses).

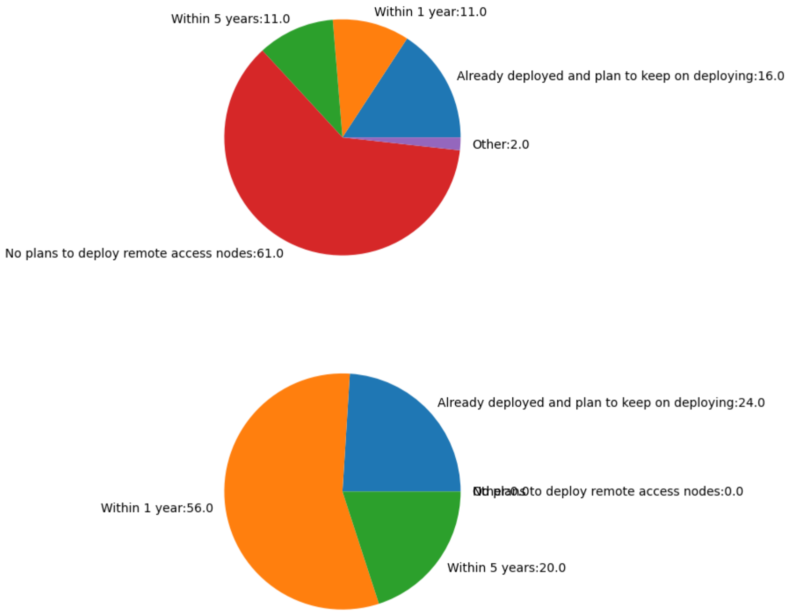


Fig. 73. Deployment of remote access nodes close to the customer (numbers show percentage of sample size, NOG on top)

For fronthaul / midhaul at macro cell sites, what type of network service have you deployed/purchased most commonly?

MPLS service dominates both samples for both fronthaul and midhaul, but dark fibre is a surprisingly strong contender in the NOG sample (Fig. 74). For the incumbents (SGA sample), MEF services are the second most adopted front- and mid-haul technology, but wireless mid-haul is equally popular in this sample. This is to be expected for incumbents, who would still be collecting return on their investment in microwave backhaul systems; such wireless front- and mid-haul is also well-known for use in rural centres distant from dense urban areas. The popularity of MEF service seems to emerge as a continuation of this class of technology's popularity as backhaul for 4G networks [338]. A MEF service is an Ethernet Virtual Connection (EVC), and it can be a point-to-point, point-to-multipoint or rooted-multipoint association between two (or more) MEF UNIs (this UNI is aligned with the T RP; see Table VII).

Note that the titular "network service" is a good example of CSPs' support of other CSPs (commonly referred to as mobile network operators, or MNOs) or of CSPs' own radio access network services. Moreover, while EVC and MPLS services are competitive, they also may be complementary; for example, an EVC may use MPLS for its transport (transport is recursive). This theme will be dealt with in the analysis.

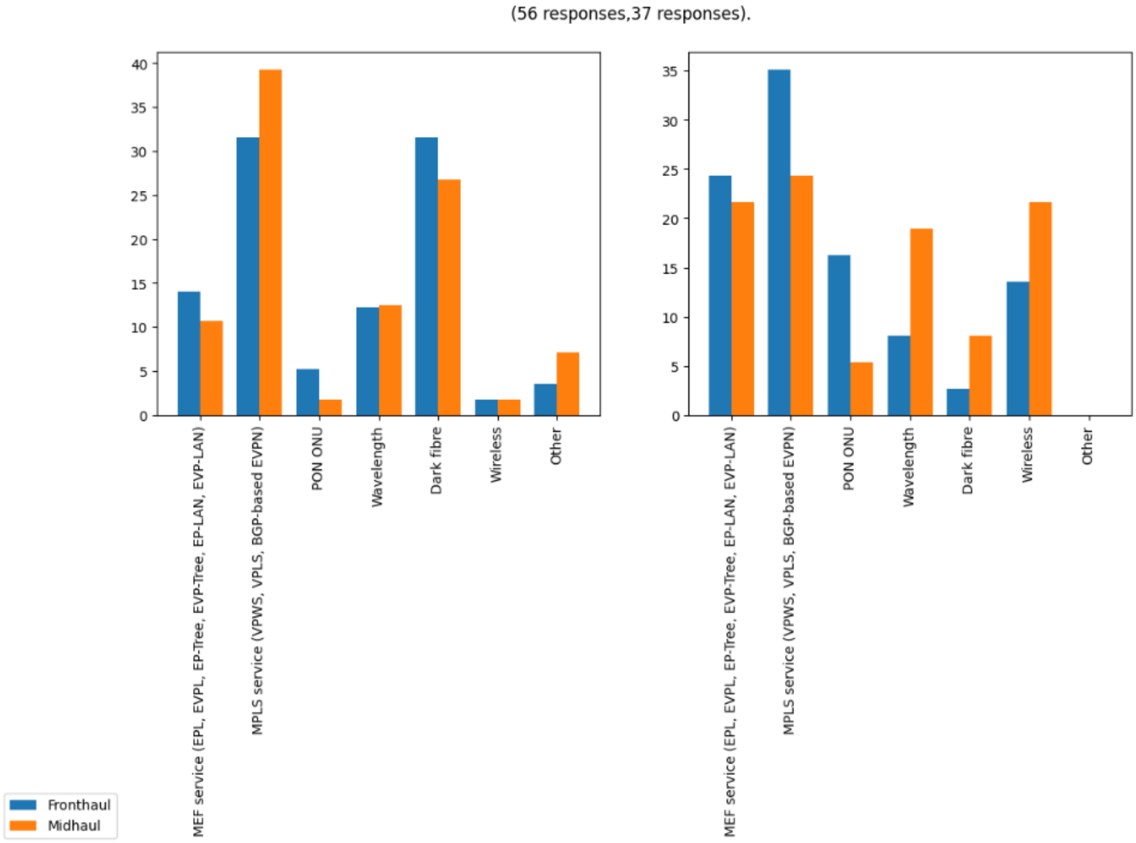


Fig. 74. Choice of fronthaul and midhaul technology for macro cell sites (numbers show percentage of sample size, NOG on left)

For fronthaul at small cell sites/fixed wireless access (FWA), what type of network service have you deployed most commonly?

A marked difference between the two profiles emerges here. Respondents from the NOG sample keep preferring MPLS service, but the incumbents lean towards EVCs, with MPLS second. Jointly, these account for 80% of the choices expressed. NOG respondents' preferences are more evenly distributed, with dark fibre accounting for 27% of the choice. Technically, this is a sound decision; fronthaul is latency-sensitive and dark fibre supports the broadest range of applications, including the URLLC (ultra-reliable, low-latency communications) swathe of the 5G application space. EVCs and wavelength services are third and fourth most common. Wavelength service consists of purchase of one (or more) wavelengths on a CSP's CWDM / DWDM (coarse / dense wavelength division multiplexing) system. If the CSP purchasing the service demands that the wavelength not pass through any active device but only through passive devices, such as multiplexers/demultiplexers (mux/demux) and ROADMs, then the latency of this service should be no worse than that of a dark fibre with equal lightpath length. With MPLS and EVC, latency constraints are more severe and CSPs' choices in this regard indicate poor interest in provision of URLLC applications.

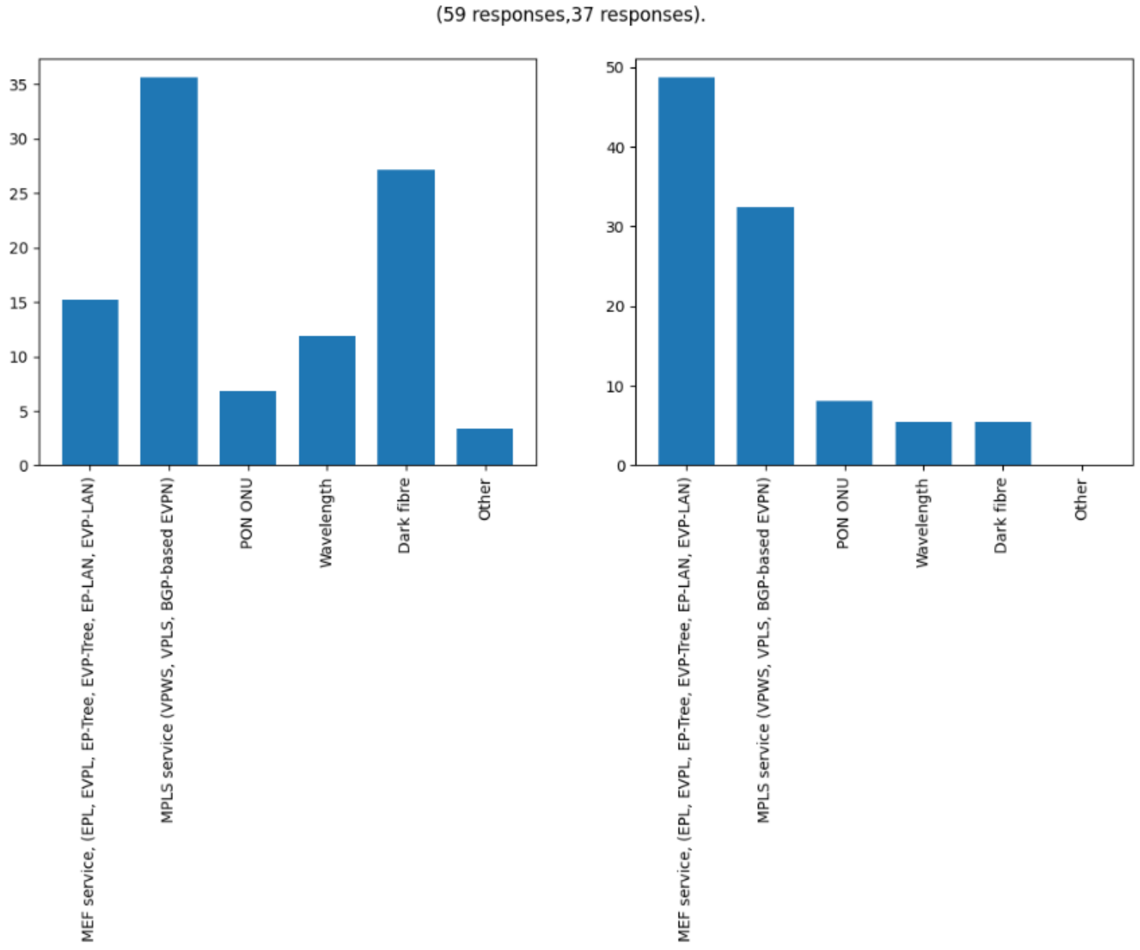


Fig. 75. Choice of fronthaul technology for small cell sites and FWA (numbers show percentage of sample size, NOG on left)

In your role as a carrier (if applicable), have you deployed disaggregated cell-site gateways (DCSGs)?

In essence, the DCSG's key attribute is that it enables separation of hardware and software on the cell-site device deployed for front-/mid-/back-haul of traffic from the cell site. Therefore, DCSGs support CSPs' need for leverage over their traditional suppliers of cell-site routing functionality. These devices are typically equipped with Ethernet or Ethernet + Synchronous Ethernet at layer 2 and layer 1, and IP/MPLS above layer 2. DCSGs are particularly relevant to CSPs who own or support other CSPs' radio access networks.

NOG sample respondents show little inclination to deploy DCSGs, while the incumbents are heavily inclined to do so. This, too, may reflect the average size of the individual respondents from the two samples.

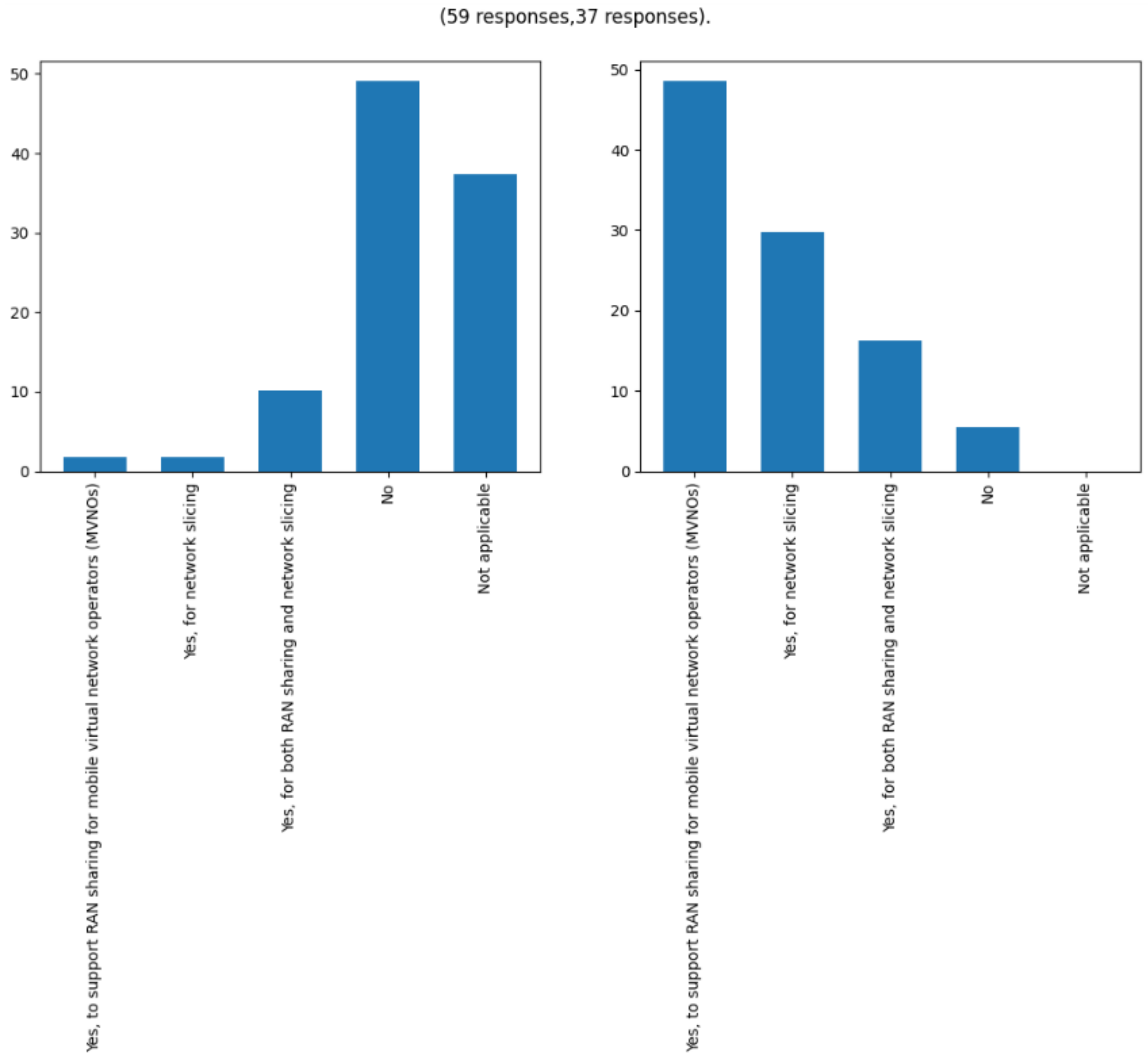


Fig. 76. Decisions of deployment of DCSGs (numbers show percentage of sample size, NOG on left)

8.3.3 Aggregation architecture: layer 2 and higher

At present, which form of layer 2 (or greater) aggregation of customer traffic from access node (V reference point) to service edges dominates?

Seamless MPLS refers to the facility to establish label-switched paths (LSPs) across segment demarcation points and CSP demarcation points. This is predicated upon the ability to exchange MPLS labels without the support of a domain-encompassing IGP. IGPs have two significant scaling impedances: one is the size of the IGP domain (the larger, the greater the scope for transient conditions) and the other is the autonomy of the individual CSPs. The ability to exchange labels is implemented in BGP-LU [339].

Fig. 77 shows that seamless MPLS is the preferred aggregation technology among NOG respondents, with provider bridging [340] second and about half as popular. On the other hand, provider bridging [340] is more popular among incumbents, with seamless MPLS second. These choices are discussed in the analysis. Fig. 78 shows that respondents' choices do not vary significantly as regards what they would do now and in future implementations (follow-up question was: For aggregation of customer traffic from access node (V reference point) to service edges, which form would you tend to prefer for current and future deployments?).

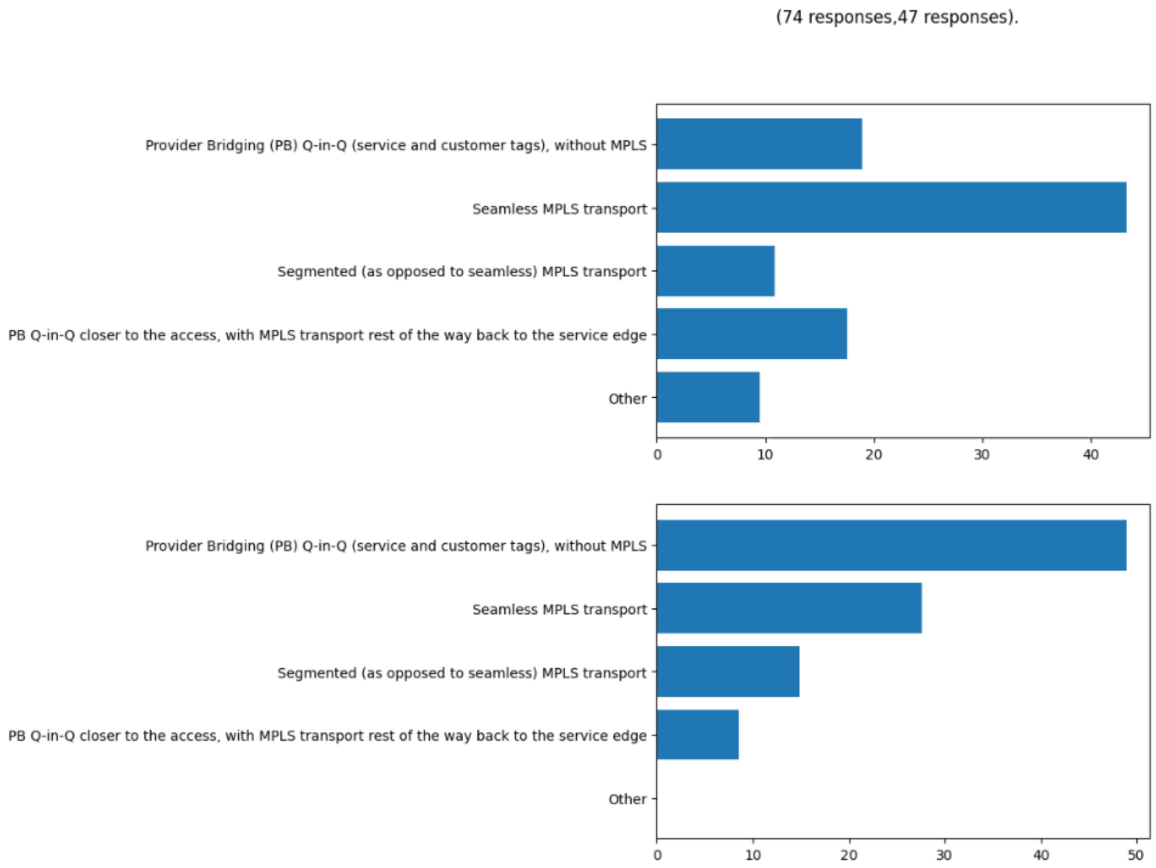


Fig. 77. **Layer 2 or greater aggregation AN – service edge (numbers show percentage of sample size, NOG on top)**

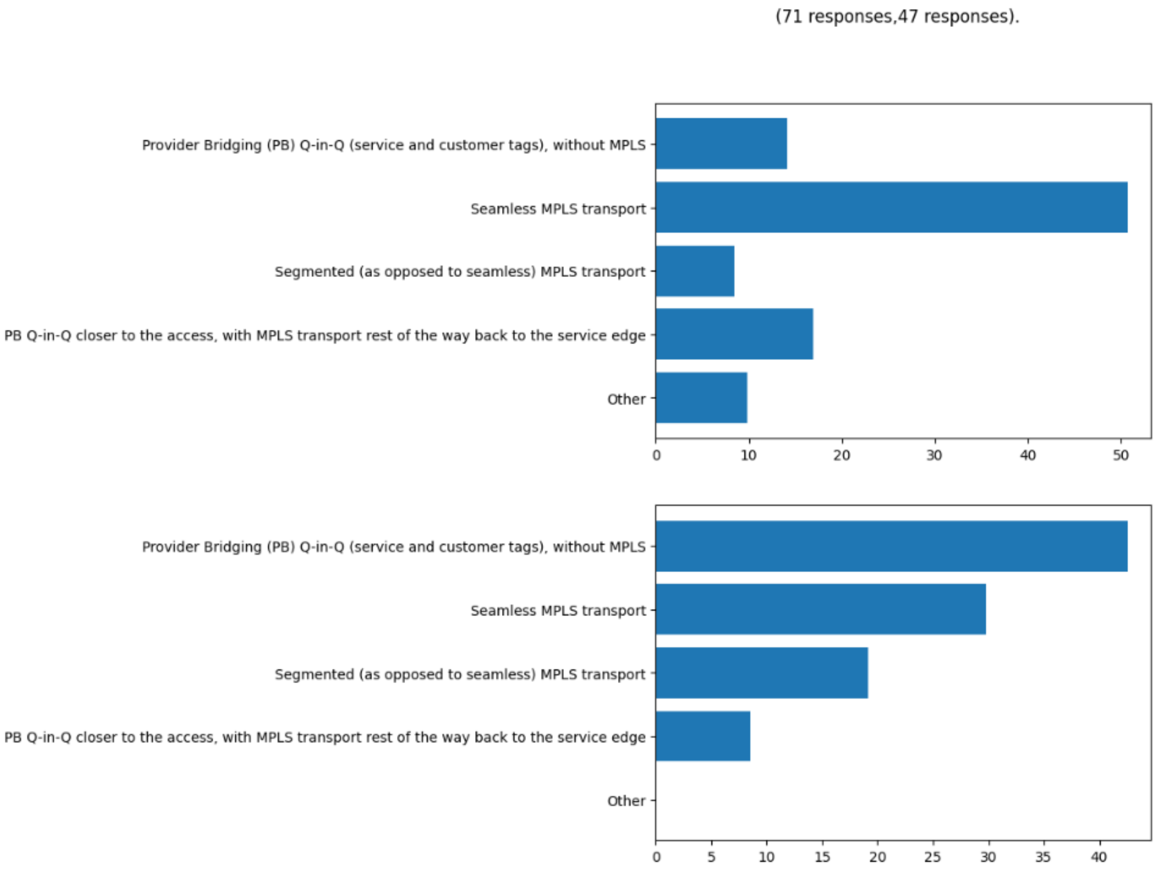


Fig. 78. **Future-oriented L2 or greater aggregation AN – service edge (numbers show percentage of sample size, NOG on top)**

Do you support the Ethernet Service Layer between the U1 and A10 reference points?

The MEF defines a three-layer Carrier Ethernet architecture (see [111, Fig. 2]), which focuses upon the Ethernet Services Layer (ETH) as the homogenizing transport across all customer access sites. Both samples clearly assert the necessity to support customer edge – provider edge connectivity over the Ethernet Services Layer (Fig. 79).

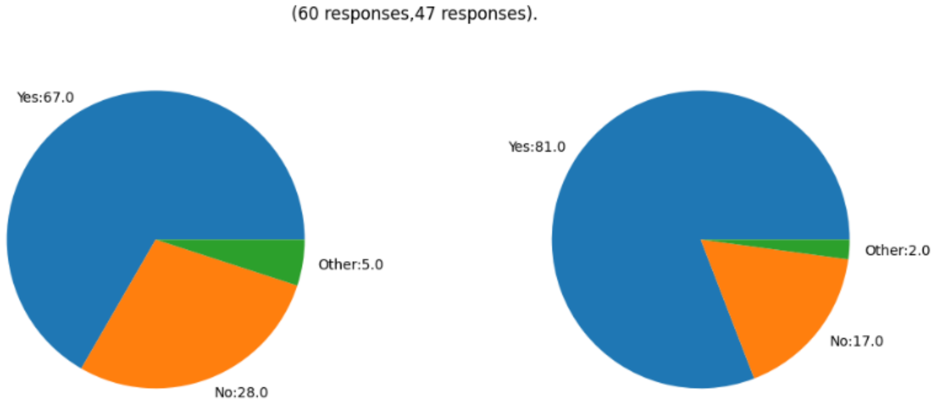


Fig. 79. **Support for Ethernet Services Layer from CE to PE (numbers show percentage of sample size, NOG on left)**

If you answered yes to the previous question, is the Ethernet Service Layer your preferred means of layer 2 aggregation?

This question was placed with the support of a diagram from BBF TR-145 [134, Fig. 5]. It asks whether the CSP prefers offering the Ethernet Service Layer to customers (note that ETH is bounded by the UNI at one end) over other L2 aggregation technologies.

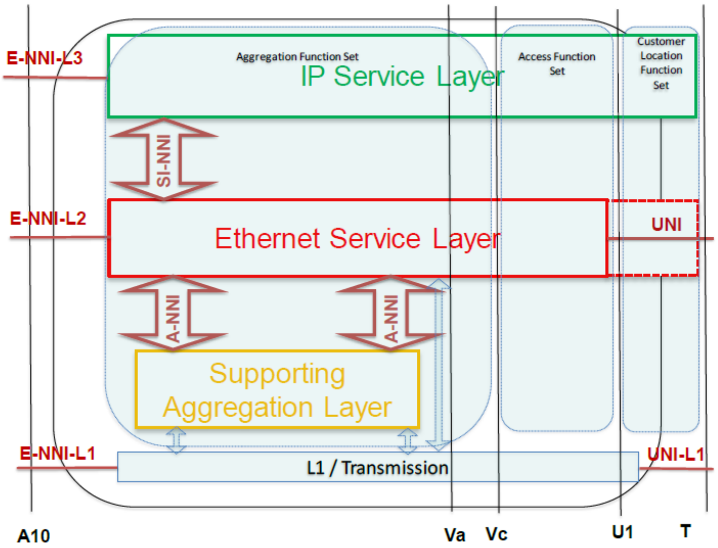


Fig. 80. **BBF TR-145's detailed reference model of multi-service broadband network (MSBN) [134, Fig. 5]**

Fig. 81 suggests that, while the layer 2 Ethernet frame is not the universal transport mechanism, it is nonetheless a dominant one. This can be understood using two interpretive keys.

1. The Ethernet frame is universally adopted in enterprise local area networks.
2. Carrier networks are increasingly more equipped with means to implement Ethernet transport, whether it be VPWS (virtual private wire service), VPLS (virtual private LAN service), or the more recent Ethernet VPN (EVPN, [341]).

These two conditions, along with this result, suggest that ETH is on its way to universality as customer-facing L2 aggregation.

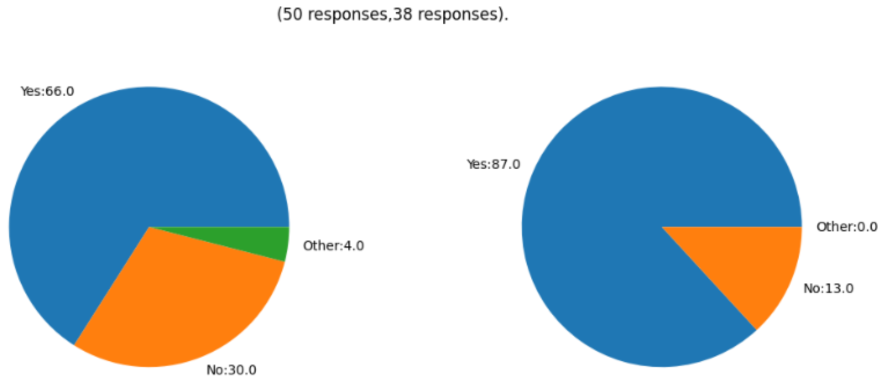


Fig. 81. **Preference for the Ethernet Service Layer as the means of aggregation (numbers show percentage of sample size, NOG on left)**

8.3.4 Aggregation architecture: layer 1 and layer 0

The following statements describe motivation for migration towards transport systems with integrated DWDM pluggable optics (and away from separate transponder/muxponder devices) and open optical line systems (and away from proprietary systems). For each motivation stated below, choose one response that best describes your opinion on its relevance as a motive for migration.

Respondents were asked to rate a series of motives (one motive per row) for moving:

1. away from DWDM systems that use separate transponder and muxponder units, and towards DWDM systems that use pluggable transceivers generating ITU-T – compliant wavelengths (coloured pluggables);
2. away from proprietary, book-ended optical line systems (OLSs), towards open OLSs.

The grid of choices is presented first (in tabular form), followed by a brief explanation of the options, followed by the results.

Table XVIII MOTIVES FOR CHANGE IN OPTICAL NETWORKS AND THEIR RELEVANCE

	1[a]	2[b]	3[c]	4[d]
DWDM optics can now be packed into switching and routing infrastructure face plates with the same density as client (grey) optics.				
A line card can now carry a mix of grey optics and DWDM optics				
400ZR and 400ZR+ standardize the physical layer for metro area networks.				
Open line systems facilitate use of interoperable pluggable DWDM transceivers.				
Open line systems facilitate integration with existing management platforms.				

[a] Mostly irrelevant

[b] Somewhat irrelevant

[c] Somewhat relevant

[d] Highly relevant

The suggestions seek clarity on relevance (as a driving motive), as follows:

1. **Density of DWDM pluggables**, in terms of the number of pluggable transceivers that fit into a single rack unit (RU). This characteristic is one of the four objectives for improvement of pluggables, i.e., cost, space (density), reach and power consumption. The latest digital, coherent, optical transceivers are now manufactured in the QSFP-DD and OSFP packages, both of which enable the packing of 36 transceivers into 1 RU.
2. **Mix grey and coloured transceivers on a line card**: this facility supports flexible upgrades from transponder-based optical networks to integrated-pluggables optical networks.
3. **Standardization** – is this an important factor in moving towards pluggables and open OLS?
4. **Open OLS is expected to support interoperable pluggable DWDM transceivers** from different vendors at both ends of the link; is this an important driver towards open OLS?

5. Open OLS is expected to include open APIs – and therefore **integrate into the CSP's choice of vendor of network management system (NMS)**; is this an important driver towards open OLS?

The results are presented as clustered bar charts. Each colour represents a specific motive, as shown in the legend below the figure. The charts are most useful if the (Fisher-Pearson coefficient of) ***skewness*** of the individual sets of bars is considered. Take DWDM pluggable density (blue bars); this chart leans the most heavily towards the right for the NOG sample. Therefore, overall, NOG respondents consider this to be the most enticing motive to migrate towards integrated pluggables, i.e., that these pluggables can now be densely packed into router faceplates. On the opposite end of the skew range, standardization is the least enticing for the NOG sample. Table XIX shows the order of relevance for both samples, basing on skewness.

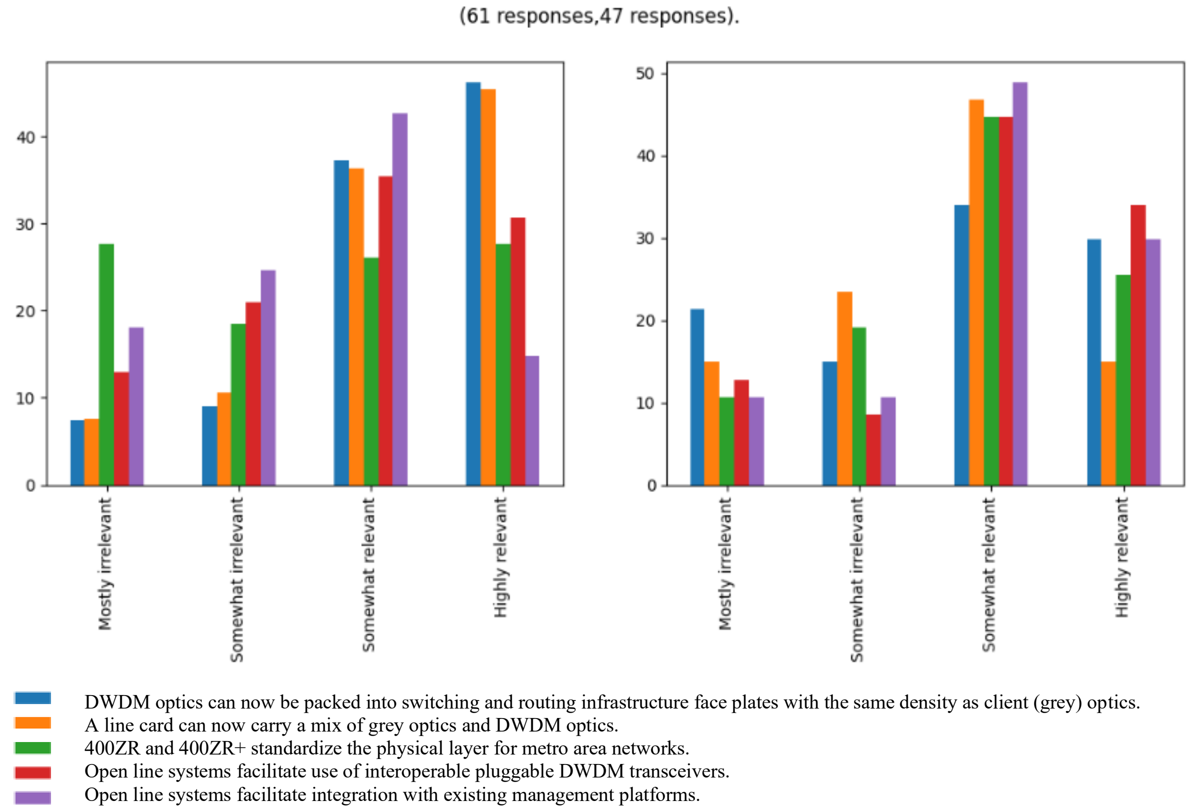


Fig. 82. Motives for migrations towards pluggable coloured transceivers and open OLS (numbers show percentage of sample size, NOG on left)

Table XIX RELEVANCE OF SUGGESTED MOTIVES, IN DESCENDING ORDER

NOG sample	SGA sample
DWDM optics density	OLSs & interoperable pluggable DWDM transceivers
Line card can mix grey and coloured transceivers	OLSs & management platforms
OLSs & interoperable pluggable DWDM transceivers	400ZR and 400ZR+ standardize MAN physical layer
OLSs & management platforms	DWDM optics density
400ZR and 400ZR+ standardize MAN physical layer	Line card can mix grey and coloured transceivers

XR optics enable a new point-to-multipoint network architecture. Do you plan to deploy this technology in your metro aggregation network?

“XR optics” is the catch-phrase for identification of an optical network technology that divides a wave band into a number of sub-bands, and supports aggregation of these sub-bands to fit different traffic distributions. This support is complemented by the facility to aggregate the sub-bands onto a single transceiver, thereby achieving a point-to-multipoint relationship between upstream and downstream transceivers.

The polarized positions of the two samples – the NOG respondents broadly do not plan to deploy, while all but three of the forty-seven SGA respondents are in some stage of engagement – again seems to suggest a division along scale of operation. That incumbents have greater traffic volume and therefore greater scope for exploiting the technology’s economies of scale, is reasonable and seems to emerge from the case studies (described later).

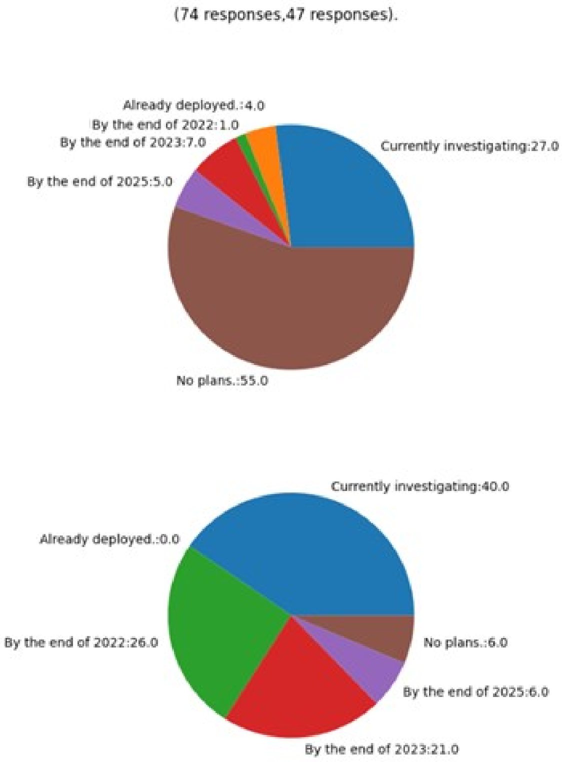


Fig. 83. Plans to deploy XR technology (numbers show percentage of sample size, NOG on top)

Claim: "XR optics point-to-multipoint network architecture will replace all other network architectures in metro aggregation".

The purpose of this question was to gather feedback from CSPs about a technology that is relevant to two (2(a) and 2(b)) of the three pressures in the second major axis of the development framework.

NOG sample respondents are largely dismissive: counting lack of consideration, and lack of agreement on its potential to replace all other network architectures, just over two-thirds of respondents do not consider it to be a dominant technology. This fraction of the sample reflects the 55% that have no plans to deploy the technology.

SGA sample respondents, on the other hand, are overwhelmingly optimistic about the technology future dominance. Almost all (95%) see this technology as more ("fully agree") or less ("somewhat agree") that XR optics' aggregation economics and granularity will displace other optical network aggregation technologies. On the basis of incumbents' scale of operations, this result suggests that the technology's implications must weigh heavily on the selection of scenarios for which to develop an implementational model.

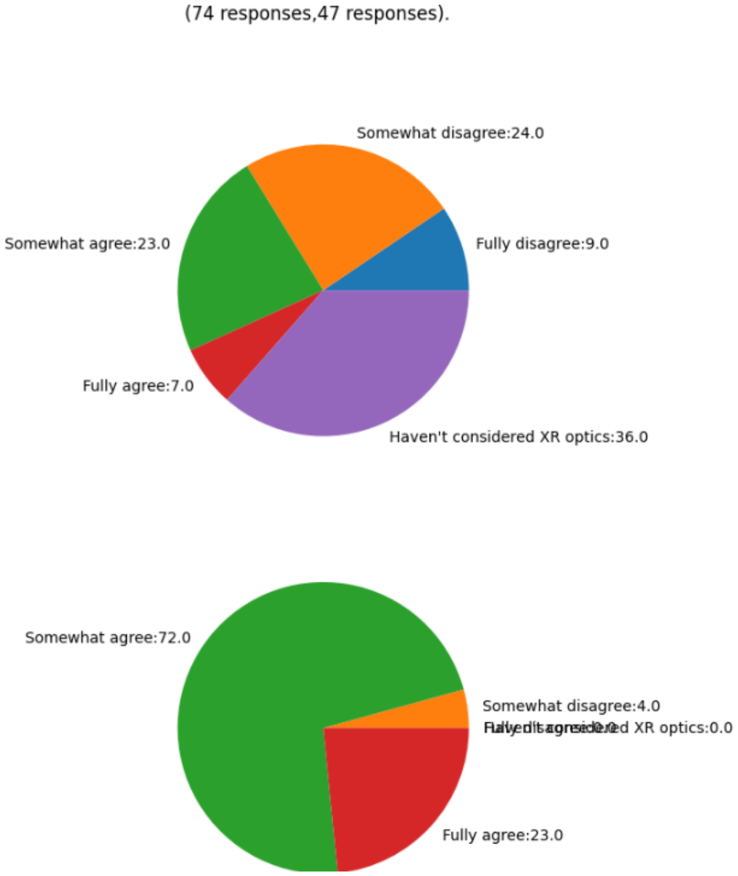


Fig. 84. Opinions on XR optics' prospect of dominating metro-aggregation (numbers show percentage of sample size, NOG on top)

Claim: "Existing OTN aggregation will stay in my network but I won't choose OTN for any expansion of my aggregation network".

Optical transport network's (OTN's) role in aggregation has been debated by a panel of CSPs and their vendors during an online symposium ([180, N. @60:25], and again at [180, N. @81:32]). An OTN equipment vendor identified it as a technology that can meet URLLC's low-latency requirement, but a CSP identified FlexE for that role, and lamented OTN's cost: "*from the perspective of cost … definitely a no-brainer*". This anecdote is symptomatic of a broader debate over whether OTN is relevant to packet networking. OTN vendors cite physical separation (in separate OTN frames) as an advantage over purely packet networks, and cite higher efficiency (better utilization) over wavelength services. On the other side of the debate, cost, simplicity and (lower-cost) alternatives such as FlexE are cited as reasons for avoidance. However, even detractors see OTN's role in mid- and long-haul networks. There is a clear need for feedback from CSPs on their intentions with regard to OTN, and it was solicited by this question.

NOG sample respondents are inclined towards dropping OTN from their plans for the future of aggregation: 58% vs 42%. OTN sample respondents are **heavily** inclined towards dropping OTN. The overall verdict is that OTN's participation in the set of metro-area aggregation technologies is of secondary importance. The difference between the two samples might be attributed to NOG respondents' failure to respond to a question which they saw as irrelevant to them, as they do not operate OTN. In that case, the two distributions would be much closer to one another.

(57 responses,47 responses).

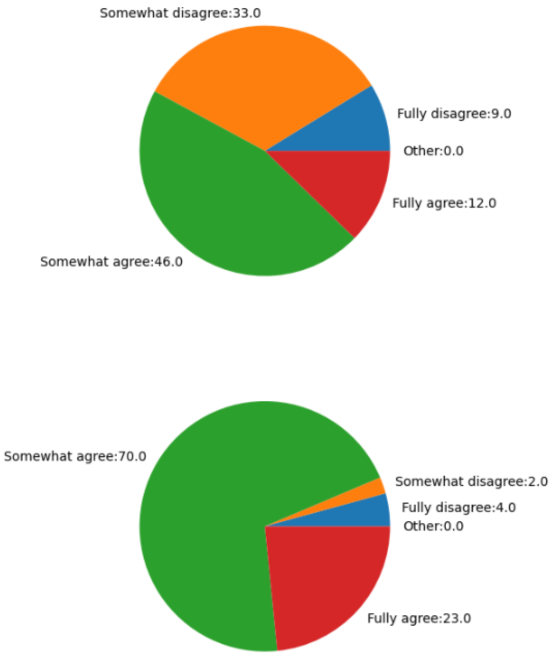


Fig. 85. Prospects for OTN's future in aggregation (numbers show percentage of sample size, NOG on top)

If you chose “somewhat agree” or “fully agree” that OTN won’t be included in expansion of your aggregation network, please indicate the reasons driving your choice.

This question follows for those who are inclined to move away from OTN. Fig. 86 shows that cost is the primary driver for both respondent samples.

The NOG sample’s respondents included three (8% of 36) explicatory comments: “*cheap fiber availability vs expensive OTN*”, “*IT (sic) is another management layer causing complexity*” and “*No Point doing TDM in todays world*”. Apart from affirming cost, these comments indicate that simplicity would have been a good addition to the list of options offered (although: a catch-all option to suggest a reason was offered).

The response turnout was low on the NOG side. This too might be attributed to the same reason as that suggested to explain the previous question’s turnout, i.e., lack of interest in a technology irrelevant to one’s own technology set.

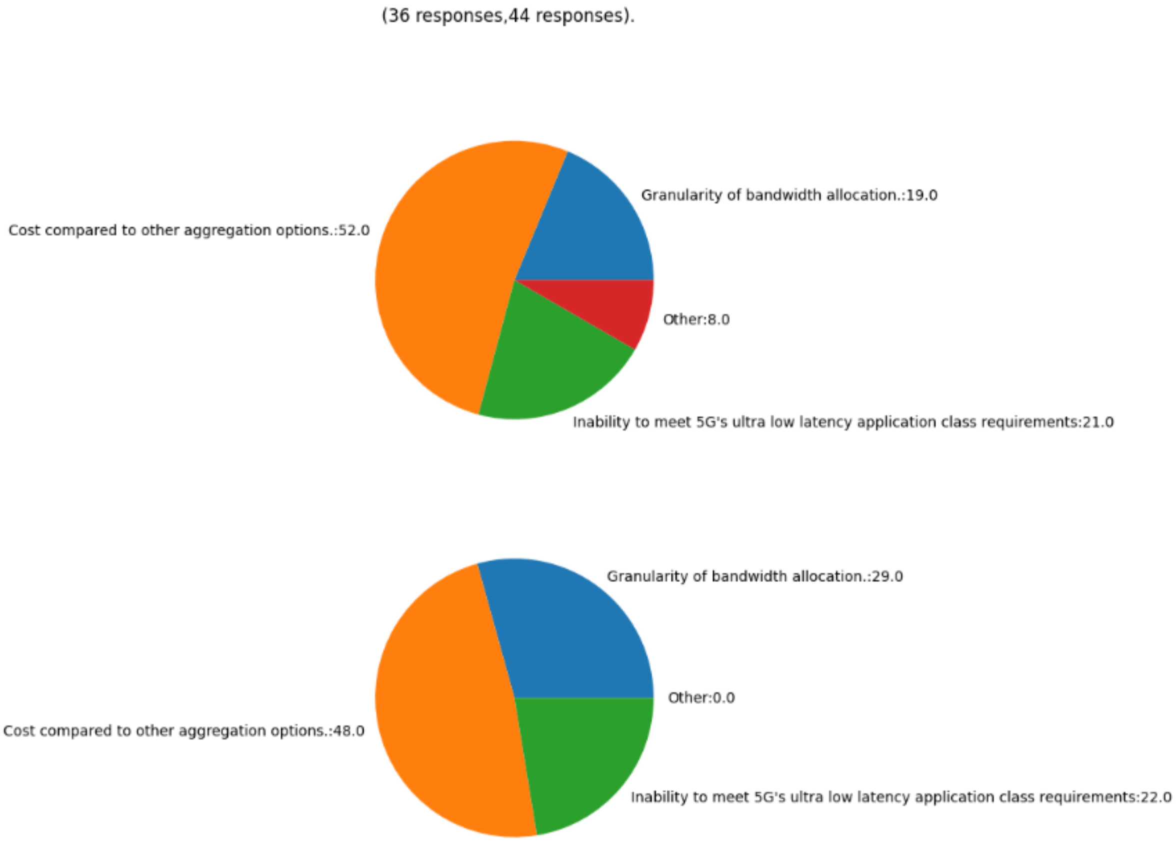


Fig. 86. Motives for dropping OTN from aggregation (numbers show percentage of sample size, NOG on top)

Claim: “Packet-based networks that share link capacities using soft slicing and/or hard slicing will fully displace OTN from metro area networks. The exception is in data-centre interconnect, where capacity allocations are stable.”

This question was intended to assess respondents’ inclination towards packet-based networks as an alternative to OTN. The basis for the exception for data-centre interconnect (DCI) comes from [342]; this source casts doubts on OTN’s prospects for dominance, but identifies it as a means to groom client traffic (crossing datacentres, hence DCI) into optical payload units (OPUs) at varying bit rates. Fig. 87 shows the results: 64% of the NOG sample and 89% of the SGA sample lean towards a fully packet-based network. One respondent in the SGA sample chose the “other” catch-all and wrote “not sure”. When this is compared with the results of the question about OTN’s future prospects in metro-aggregation, it can be seen that more NOG respondents have a long-term vision of MANs without OTN, than those who favoured dropping the technology from the aggregation set. Moreover, once again, if the unresponsive subset of the NOG sample is considered as in tacit agreement, then the NOG sample would include 72% who are inclined towards purely packet-based networks, without OTN.

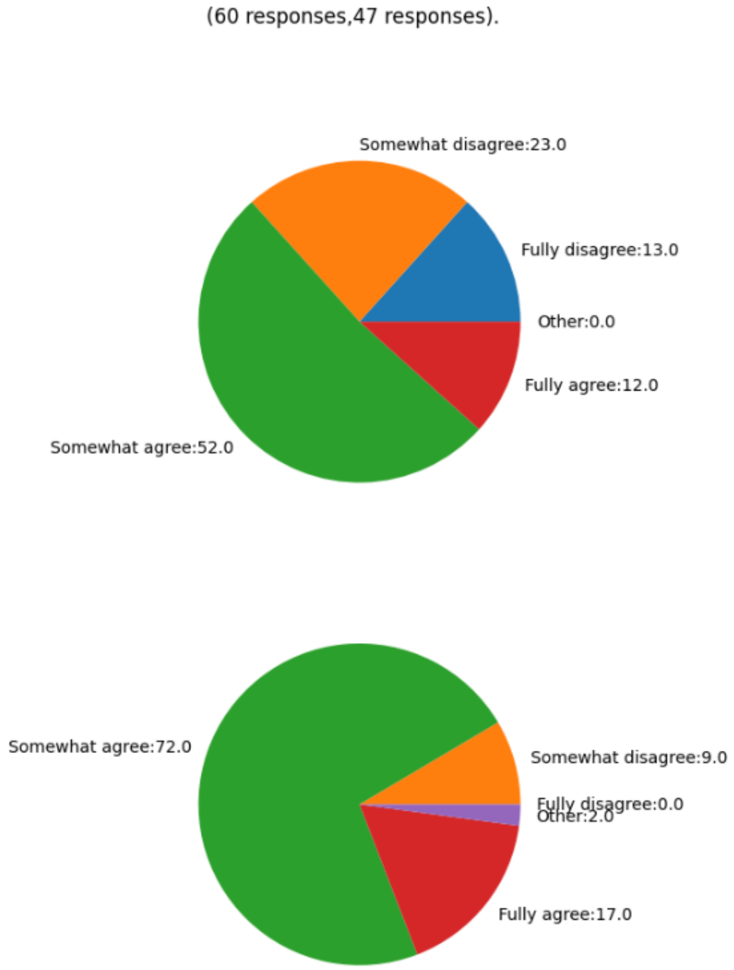


Fig. 87. **Packet-based networks will displace OTN from MANs, except in datacentre interconnect (numbers show percentage of sample size, NOG on top)**

8.3.5 Transport network architecture: stacking layers, from 0 to 3

The objective here is to detach from the details and attempt to acquire an understanding of trends in transport network architecture.

Which of the following best describes your current dominant form of metro-aggregation?

This question seeks to elicit an understanding of which technologies are occupying the stack of layers in transport, from layer 0 up to 3. Here, the qualitative analysis must be pre-empted, as otherwise the NOG sample's distribution cannot be interpreted correctly. Suffice it to state, for now, that "routed optical networks over Ethernet without ROADMs" is taken to refer to the practice of drawing fibre up to the router chassis or shelf, and plugging it in to transceivers capable of meeting the optical link's budget – if possible, without line amplifiers.

With this proviso, the results (Fig. 88) seem to be very much in accordance with the broad divide between newer entrants (NOG) and incumbents (SGA). NOG respondents prefer the shallowest, least recursive stack of layers, and few of them are still using SDH/SONET. Their second preference – essentially the addition of coloured pluggable transceivers and ROADMs to support optical bypass of a switching node – is a distant second. It also emerges as the SGA sample's respondents first choice. For the SGA sample, OTN is broadly deployed as a sub-wavelength service layer for the carriage of Ethernet frames. OTN is far less popular among the NOG sample's respondents.

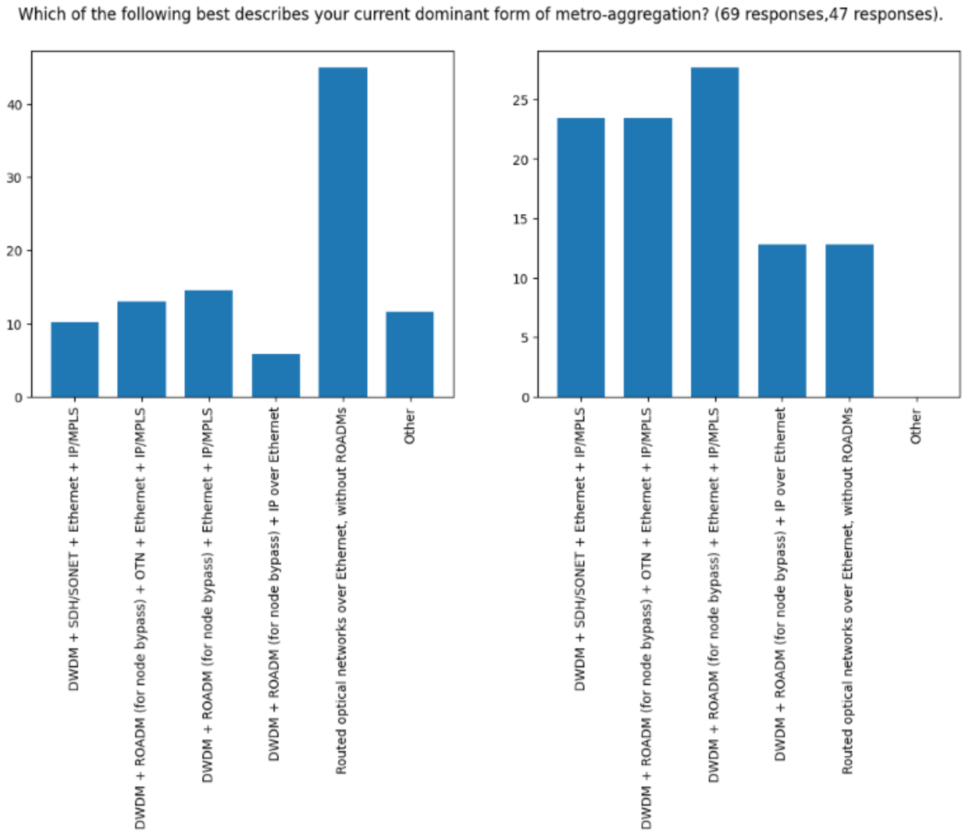

Which of the following best describes your current dominant form of metro-aggregation? (69 responses,47 responses).

Fig. 88. Current dominant form of metro-aggregation (numbers show percentage of sample size, NOG on left)

Which of the following best describes how you would deploy a greenfield form of metro-aggregation?

In order to understand which scenarios to model, it is necessary to solicit CSPs' plans for the future of the transport network. The NOG sample's respondents' answers are consistent with previous responses: OTN's presence in the stack of layers comes in a distant circa 10% of respondents' choices. The most common choice (about 43%) is simply drawing fibre up to the switching node and plugging it in. A close second (about 33%) uses coloured pluggables and ROADMs. Note that MPLS is a part of the transport stack in every bar but the middle one, which accounts for less than 10% of the total.

SGA's sample's respondents' answers pose a difficulty. Notwithstanding the aversion claimed earlier to further development of the metro-aggregation span with OTN, here the technology is included in the most-commonly chosen stack of layers. When this issue was discussed with SGA's researchers, my attention was drawn to the uncertainty in respondents' previous answer: they "somewhat" agreed that they would not expand OTN. This leaves room for partial expansion, where specific customer requirements demand OTN's characteristics. A further, clarifying observation is that, overall, 36% would include OTN; the remaining 64% would not. When situated within the full context, the choices in favour of OTN retain overall consistency. A remedial approach (for future efforts) might be to attempt to solicit clarification from respondents through hard-coded dependencies in questionnaires.

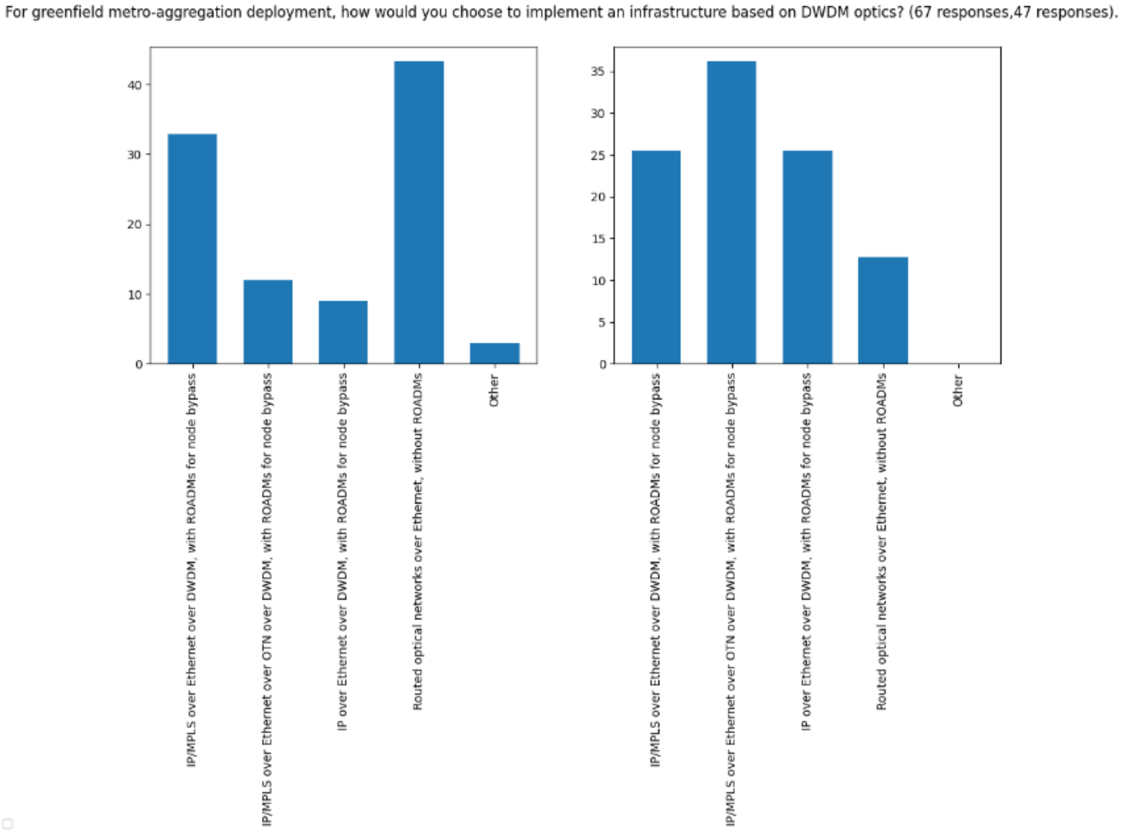

For greenfield metro-aggregation deployment, how would you choose to implement an infrastructure based on DWDM optics? (67 responses,47 responses).

Fig. 89. Desired (greenfield) form of metro-aggregation (numbers show percentage of sample size, NOG on left)

Claim: "In the future, a mesh network will likely replace the metro-core ring at least in urban area with challenging capacity and resilience requirements."

The basis of this claim lies in [342], which states that "*the main driver for moving to a metro-core mesh is that it enables the introduction of IP-over-DWDM multilayer resilience schemes, with remarkable benefits in terms of enhanced reliability and optical interface reduction*"; in turn, this claim is rooted in [343]. Both papers come from sources that bridge the academia – industry divide; such claims, therefore, are weighty and merit investigation. The enhancement in reliability referred to in [342] (above) is enabled by fast-reroute (FRR), which can exploit loop-free alternate (LFA) and topology-independent loop-free alternate (TI-LFA). Since these reliability schemes are obtained through the higher-layer visibility at layer 3 than the rudimentary visibility available at layer 0, then the reduction in optical interfaces ensues. Rather than being limited to the addition of (costly) optical interfaces and employment of optical (network layer) protection switching (OPS), the network engineer is armed with LFA and TI-LFA as a means to achieving sub-50 ms switching in the event of link failure.

Fig. 90 shows an inclination towards agreement with the claim, for both samples. Two NOG respondents ("other" slice of the pie) offered similar opinions: that such an architecture depends on (a) "geographies, costs and customer demand", (b) "the market, service area, and business case".

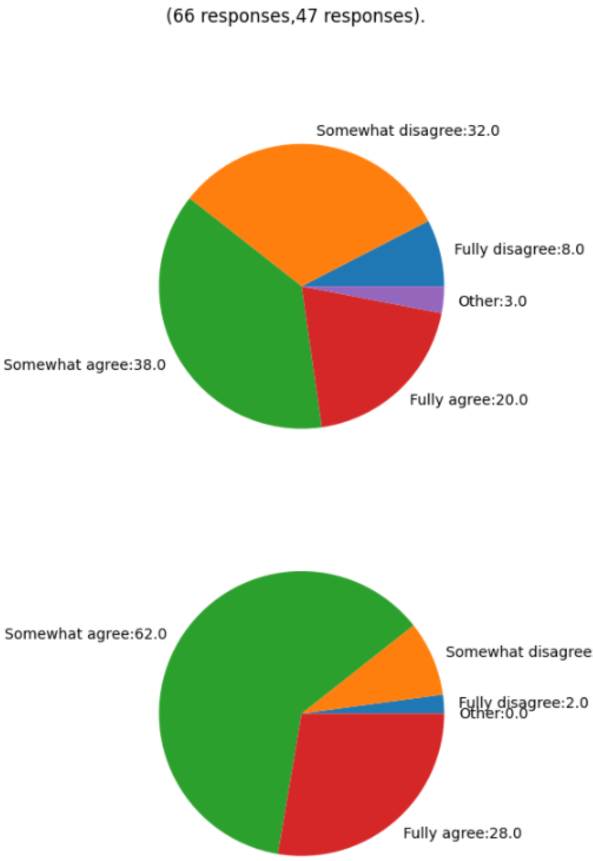


Fig. 90. CSPs' inclination towards employing meshed over ringed metro-core nodes (numbers show percentage of sample size, NOG on top)

For greenfield metro-core deployment, how would you choose to implement an infrastructure based on DWDM optics?

For NOG sample respondents, the most common choice remains the simplest one: take (dark) fibre to the switching node, colourfully referred to as “point-and-shoot” by Arelion’s (ex-Telia Carrier) representative in [182]. A partial transcript of the presentation serves to clarify the architecture represented here.

> *“We investigated every route we have between routers and found out that about a third of the routes are < 40 km & < 400G so that would mean we could replace every metro DWDM we have there with this new pluggable technology and shoot from router to router.”*

The NOG sample’s responses (Fig. 91) are consistent with the collective mindset expressed in response to the metro-aggregation questions: keep it simple and only use ROADMs if you must. The “other” case shown in the left-side chart expressed “our toplogy *(sic)* - [IP or Ethernet] over MPLS SR using BGP EVPN and coherent optics wherever required”. This seems to qualify as the majority case, i.e., routed optical networks over Ethernet without ROADMs, and observes that the control plane uses Segment Routing. The SGA sample’s responses are more evenly distributed; OTN is included in about 36% of the responses and excluded from 64%. This is consistent with previous choices on the metro-aggregation stack.

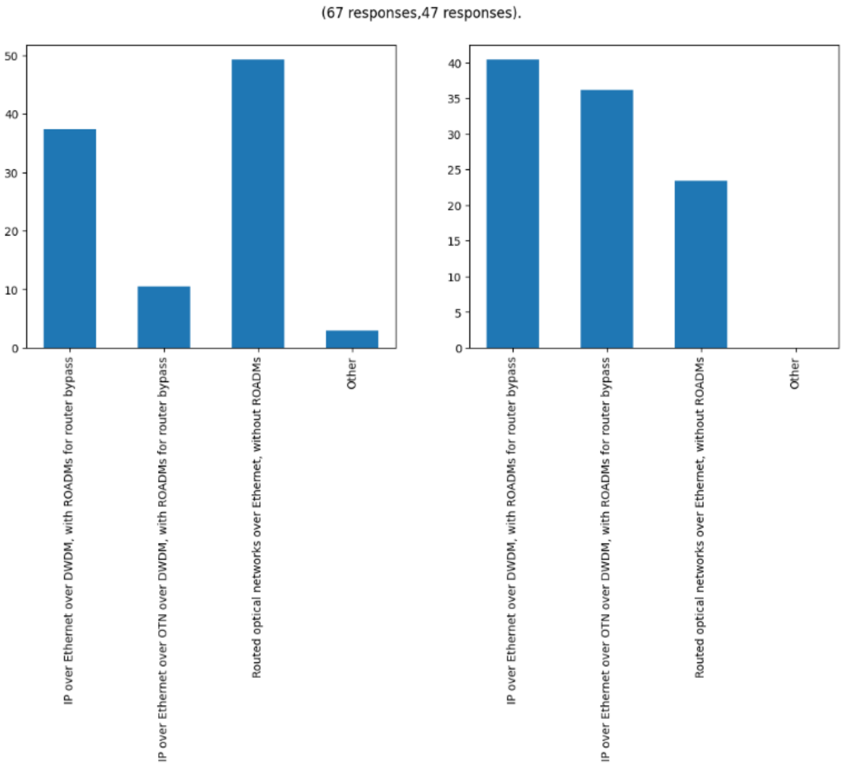


Fig. 91. Desired (greenfield) form of metro-aggregation (numbers show percentage of sample size, NOG on left)

8.3.6 Service edge

The final series of questions explored CSPs' thoughts about the development of the service edge. In particular, thoughts on locating the edge closer to the customer were sought.

Do you plan to deploy remote access nodes (Option 0) to enable MEC services? (B) Service edge locations (BBF TR-178): which do you currently employ for {Internet/Video} Broadband Network Gateway?

The two questions (in this sub-sub-section's title) complement one another and serve as an investigation of the relationship between CSPs' propensity to set up deep (close to the customer) service edges, and their disposition to deploy MEC nodes. The questions were accompanied by a graphic from the BBF's TR-178 (see Fig. 92, which is the same as Fig. 72, and reproduced here for convenience's sake).

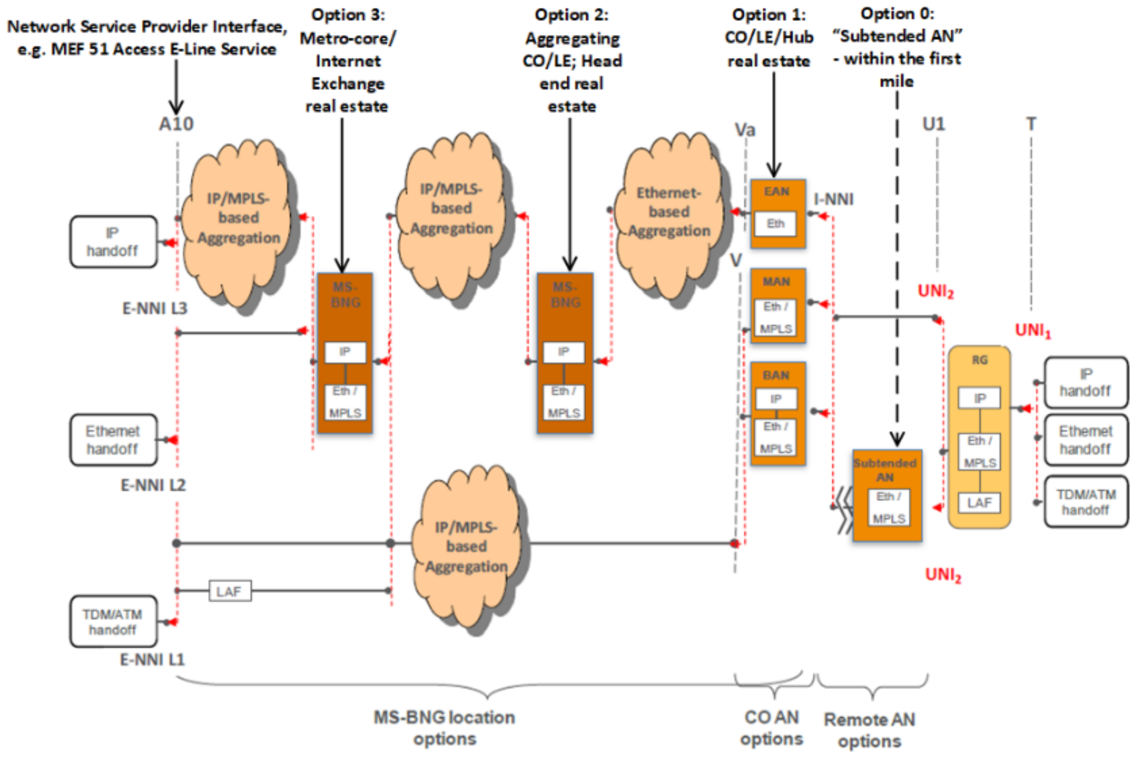


Fig. 92. General TR-178 architectural scheme, encompassing its targeted deployment scenarios [337, Fig. 2]

Deployment of remote access nodes at the location indicated by Option 0: Fig. 93 shows the distribution of the responses. In the NOG sample:

- only 57 out of a total possible of 79 answered, and in this subset, 61% (35 out of 57) have no plans to deploy.

- Nine (9) already have deployed MEC nodes at the Option-0-location. However, out of these 9, only one has also deployed an Internet BNG at this location (this was determined through cross-reference within the nine respondents' answers).
- As regards those who installed an Internet BNG at the Option-0-location, none of them plan to deploy MEC nodes.

The above observations indicate that there seems to be no correlation between the two facilities. Similarly, only one NOG sample respondent has both MEC facilities and a video BNG at the Option-0-location.

As regards the SGA sample:

- all 50 participants responded.
- 12 (24%) claim to have deployed Option 0 MEC nodes, yet none of them have an Internet BNG at this site.

Given the similarity between the two samples in the lack of correlation between deployment of BNGs and MEC at the Option-0-location, one possible interpretation might be that the facilities that have to date housed BNGs, whether for Internet or video service, do not have sufficient infrastructural provision (power, cooling and security) to support MEC hardware.

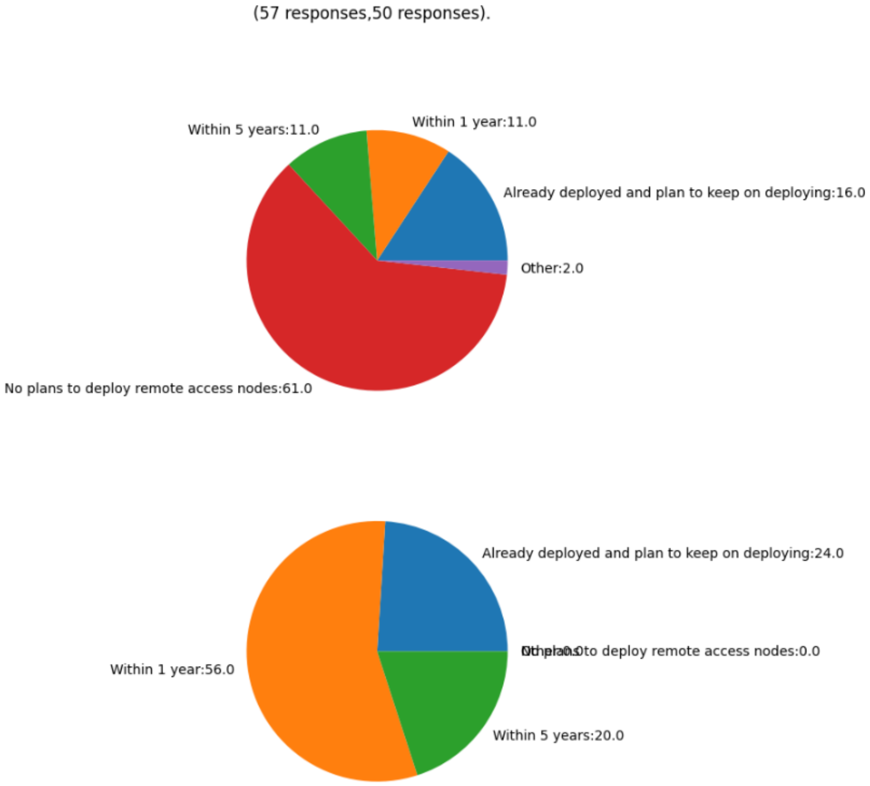


Fig. 93. CSPs' plans for deployment of remote access nodes deep into the access segment – Option 0 (numbers show percentage of sample size, NOG on top)

SGA sample respondents seem keen to move towards deep MEC nodes; indeed, as incumbents, they would have greater freedom to specialize personnel and re-purpose real estate.

One useful observation that emerges is the similarity of the distributions for the two samples, and its match to the expectation that the mode of the distribution lies at Option 2 – i.e., the CO/LE.

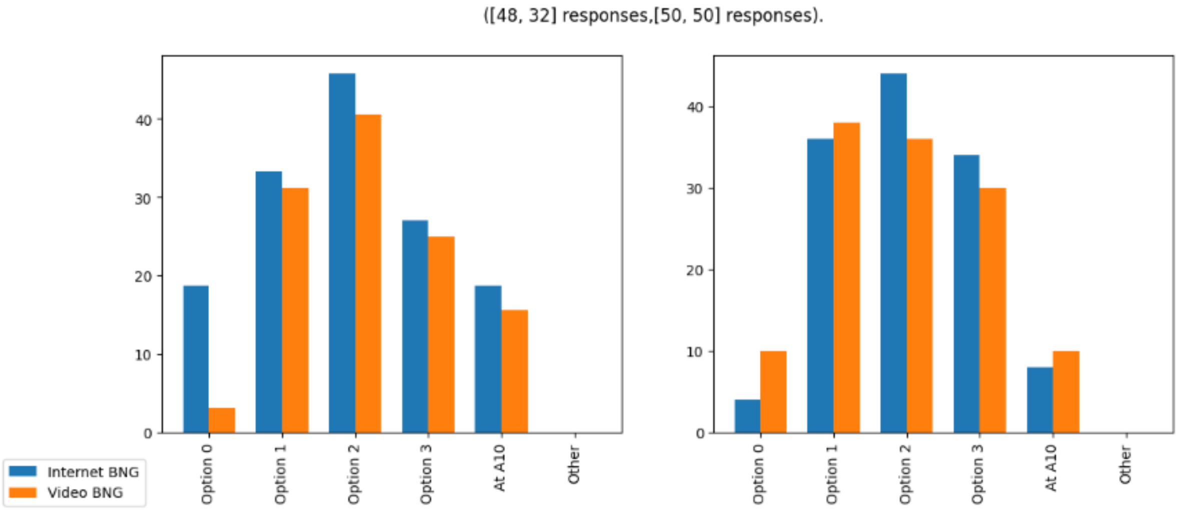


Fig. 94. Service edge locations supported by CSPs (numbers show percentage of sample size, NOG on left)

Claim: "Support for enhanced mobile broadband (eMBB) is improved by adding video BNGs closer to the end user.", and (B) Claim: "I would consider adding video BNGs closer to the end user to improve energy efficiency of video delivery."

These two questions were expected to draw a – more or less – affirmative collective response. Some of the NOG sample respondents declined to answer the question on improvement of support for eMBB, on the basis of their lack of support for a mobile network, or on their lack of support for video service. The NOG response to the claim for improved support for eMBB was further nuanced by some technical observations such as "Video is becoming more unicast, therefore it is a capacity planning equation that determines where video BNG is placed" and "Depends on distance of the BNG". Similar nuance was expressed by some NOG respondents on the issue of energy efficiency, e.g., "energy efficiency needs to be considered in its totality not only on the intefaces *(sic)*. Typically the best energy efficiency is obtained in the data centres". A general statement can be made that the NOG sample reflects an expectation that both support for eMBB as well as its energy efficiency are improved by locating video BNGs closer to the end user. The SGA sample reflects a more emphatic expectation of improved support and energy efficiency. These results suggest that CSPs are favourably inclined towards deploying video BNGs.

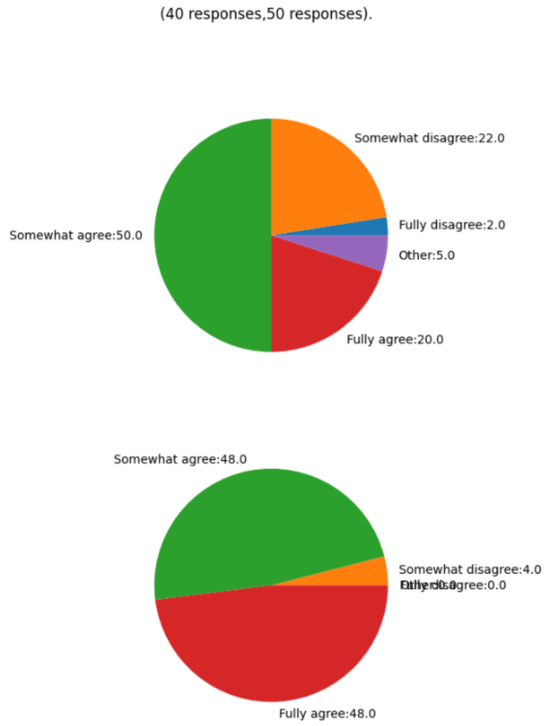


Fig. 95. CSPs' understanding of whether support for eMBB is improved by adding video BNGs closer to the end user (numbers show percentage of sample size, NOG on top)

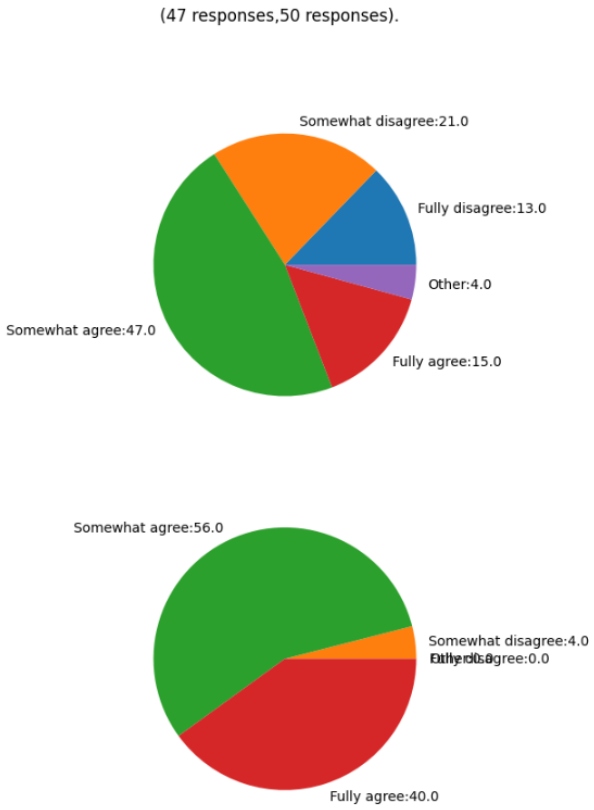


Fig. 96. CSPs' understanding of whether energy efficiency is improved by adding video BNGs closer to the end user (numbers show percentage of sample size, NOG on top)

8.4 Qualitative survey results

The two broad objectives have been identified as discussion of the graphical summaries of the results, and assessment of the objective clarity of the questionnaire's questions. I have refrained from further qualification of the objective (e.g., it might seem plausible to express the objective as a desire to learn about reviewers' opinions on its accuracy), because the techniques (interviews, emails, reports) at the foundation of qualitative surveys have the potential to supply raw data that exceeds, or at least encompasses, narrower objectives. This liberal mindset is not a substitute for accurate analysis; rather, it is a pre-requisite. Analytical techniques that develop results from qualitative surveys' raw data have been used in chapter 7, to high yield. There, the PAD method (a development of the well-known technique of structural coding), was shown to elicit trends as it worked through a corpus of research units (a collection of papers addressing a problem domain). Through three rich collections of codes, under problem, approach and development categories, it was possible to observe the formation of patterns in the linkages between subsets of each of the three categories. Therefore, while an initial statement of objectives must be broad, the processing of the raw data obtains an analysis that mines and extracts detail, including – but not limited to – the aforementioned better understanding of the accuracy of the data collected.

All the resources mined for this qualitative survey are available online. The resources have already been referred to earlier in this chapter. Table XX collects references to, and descriptions of the resources under one structure, for convenience's sake.

Table XX RESOURCES COLLECTED DURING DATA STAGE OF QUALITATIVE SURVEY

Resource description	**Ref. #**	**URL**
Face-to-face interview: Dave Eilert[45]	[333]	https://drive.google.com/open?id=1ayHhTdpfdIWMZuyLPTapKDZS1-cUFIRM&usp=drive_fs
Face-to-face interview: Ovidiu-Mădălin Roșet[46]	[327]	https://drive.google.com/open?id=1H8SeNQ7PZ5EVQ7q1fYgT4EGTYZ51mgvu&usp=drive_fs
Face-to-face interview: Haider Khalid[47]	[328]	https://drive.google.com/open?id=1LnfI6L4elnFCuz85bn6Thjwc77OYb_ji&usp=drive_fs
E-mail thread: Mark Tinka[48]	[334]	https://drive.google.com/file/d/1eltC-QaLPiZitrlngEWAUS5gtiimY-jw/view?usp=sharing
E-mail thread: Philip Smith[49]	[323]	https://drive.google.com/open?id=1bN0RS_M93UojmdA0oYMYqJI9Ql_uM8Hk&usp=drive_fs

---

[45] https://www.linkedin.com/in/dave-eilert-3a1a17b/
[46] https://www.linkedin.com/in/ovi12/
[47] https://www.linkedin.com/in/haider-khalid/
[48] https://www.linkedin.com/in/mark-tinka-5b03055/
[49] https://www.linkedin.com/in/philip-smith-154502/

E-mail thread: Daniel King[50]	[332]	https://drive.google.com/open?id=1ayHhTdpfdIWMZuyLPTapKDZS1-cUFIRM&usp=drive_fs
E-mail thread: anonymous	[318]	https://drive.google.com/open?id=1el7oH18k_dWsupVPNjCfGHSAaa_7tkX3&usp=drive_fs
Written assessment: Haider Khalid[51]	[331]	https://drive.google.com/file/d/1gjth5v8W2x4Dll_l5oyPW9n39uT9fnGc/view?usp=sharing

8.4.1 Face-to-face interview saliencies

Face-to-face interviews were conducted online, over Zoom, and recorded with the permission of the interviewees. Duration ranged between 50 and 90 minutes. Dave Eilert and Haider Khalid were recruited by SG Analytics, on the basis of the anonymous reviewer's advice, with which Mark Tinka concurred, to source interviewees directly involved with daily network operations. I recruited Ovidiu-Mădălin Roșeț directly, on the basis of the same advice and on the basis of personal experience of his technical networking skills. Both Haider and Dave are CCIEs.

Face-to-face discussions provided the opportunity to listen to narratives that described metro-aggregation and metro-core architectures. Such narratives address the degree of assurance which can be obtained about the integrity of the communication channel between questioner and respondent in the impersonal medium of the questionnaire. That is: since questions are text on paper, they are always, to some degree, subject to interpretation. The medium of a discussion reduces the subjectivity; the parties in a discussion have reasonable opportunity to solicit clarification in case of doubt. Indeed, the responses to the questionnaire ***did*** include the occasional unexpected interpretation. For example, one otherwise coherent respondent claimed, when asked about the location of the Internet BNG, that "BNGs have been dead for a long time! We use IP only, termination happens directly on the connected switch unless I wholesale circuit in which case its dragged over to the handover". Here, it seems that the respondent was thinking about the service edge in terms of the PE router. This is indeed a possible interpretation: it is the "IP edge" that was referred to in this work in chapter 4; what exactly has "been dead for a long time" is unclear.

---

[50] https://www.linkedin.com/in/danielking/

[51] https://www.linkedin.com/in/haider-khalid/

“the old way of doing things”

In the following, references to the CSP in the object of the discussion are replaced by the placeholder CSP_X. Suffice it to say that CSP_X is a highly reputable organization, and is not an agent of dubious practices.

“So from what I’ve seen, and now, I’m talking about some providers, …, more about providers in Romania and some [] in the region.

So it’s [] a thing of evolution and maybe that’s also a thing, I don’t know. What’s the budget of those companies and how long has it been since they’ve upgraded and so on.

Because … ten years back you would have a lot of switches in the access and aggregation, because the switches have high ports density. They’re quite cheap. And you could deploy either spanning tree, for redundancy items, you know, those ring topologies where you deploy them for the loop prevention.

And this is where this PB would apply if you would have a pure layer 2 network in the access and the aggregation, … you would configure the IP … the service layer would be on some routers on the core and you would have pure layer 2 on access and aggregation switches.

But the PB, I think it is needed … if you have many, many customers, then you would start needing different VLANs – S-VLANs, C-VLANs and stuff like that.

In CSP_X in Romania, so I don’t know exactly how many customers it had. And also I have to mention something else, first of all, CSP_X didn’t provide broadband services up until just a few years ago.

So what I am talking about is CSP_X network, while they had sites of telco, so 2G, 3G, 4G and so on, and enterprise customers, but not residential. So we didn’t have BNG, we didn’t provide those services and that’s what I am referring to.

And when I first started working at CSP_X, we had what I was saying: a lot of switches and IP on some core routers and you would just have VLANs for each of your customers.

If maybe we had more than 4000 customers, we would run out of VLANs, we would deploy C-VLAN, service VLAN and so on, and maybe got into PB and Q-in-Q. But in our case, it wasn’t. It wasn’t needed.

So maybe the size is also important because you’d need to fit more services and you would need more VLANs and you would deploy PB.

But … this is the old way of doing things. No one sane would deploy now switches in access and aggregation and no integrators, vendors are offering this solution. So: I also work for a system integrator and I wouldn’t go with switches in the access or aggregation.”

This assertion begged a clarification: what features would one expect on these Ethernet switches? I posed this question.

"The basic ones so Q-in-Q would be one of them, just to carry VLANs and have some kind of loop-prevention protocol, like as I said, spanning tree, so really basic. I don't know how people in other countries say it, but here in Romania, we say just to carry those VLANs."

Key takeaway: Ethernet switches supporting spanning-tree protocol and basic VLANs were the cheap and effective way of aggregating from the access node upwards towards the provider edge of the metro core.

"the intermedia[][te] step"

I pressed for clarification on what the future of metro-aggregation looks like. Ovidiu's position as a system integrator enables him to broaden his vision beyond an extant portfolio of technologies, towards accommodation of a diverse set of customers, whether CSPs or otherwise. The answer branched into validation of the NOG respondents' commonest choice of layer 2+[52] aggregation (seamless MPLS), and then branched again towards a welcome insight into how large routing domains can be constructed.

"So I'll tell you what we're moving into, but I'll give you the intermediary (sic) step ... in CSP_X we moved from layer two switches, from Ethernet switches in access and aggregation to bringing the IP into the access area. And that involved seamless MPLS. You could deploy one big domain, let's say one IP domain one, let's say ISIS routing domain and put inside that routing domain our whole routers in Romania.

But you have ... do you know why you can't actually do that in large networks because if not, I can explain."

I certainly was not ready to claim that I knew all that Ovidiu knows, and gladly accepted his offer to elaborate.

"*So you would have an IGP that needs to run the shortest path first algorithm on 1000 nodes. If you have a small change in one part of your network, that small change or failure would cause an SPF re-run on a totally different part of your network and it's not viable. And you would need a lot of resources on access nodes, because you would need to have the whole map. And ... this can't be done.*"

It was next explained how seamless MPLS solves this problem, and plays an important role in the stack of transport layers.

---

[52] Layer 2+ is a somewhat loose term that is used here to refer to technologies that fit above OSI layer 1 but below OSI layer 3. In this work, the only significant members of the implied set are IEEE 802, 802.1, 802.3 and MPLS (RFC 3031).

"This is where seamless MPLS comes into play because you would just split. Let's say you have Romania and you split it in ten regions and each region is has its own routing domain and you still need to provide end-to-end LSPs, and you use BGP labelled unicast for that, to advertise the loopbacks for other regions, to advertise the label for them. In this way you use BGP labelled unicast, and this is the next step that providers move to from layer two to the three in the access."

The next statement was an assertion of first-hand source status:

"That's when I got in CSP_X, when we were migrating from that old network with Ethernet switches to this new network with seamless. And I'm quite aware of what we were doing and how we were moving services from the core to the access. And I understand the reasoning and that's what I'm trying to explain and I hope it was delivered for you."

Key takeaway: Seamless MPLS is enabling "IPfication" of the network, wherein the CSP is able to bring IP connectivity into the access network. Moreover, seamless MPLS, through BGP-LU, supports end-to-end LSPs in large networks.

The future

My line of questioning moved towards provision of MEF-compliant ETH layer services, which, operators claim to support in their overwhelming majority (see Fig. 79). In the process of answering, Ovidiu started addressing the future of metro-aggregation.

"Yes, definitely. Going back to what we discussed about pure layer 2 Ethernet networks and Seamless MPLS - the next step is with Segment Routing and EVPN. That's what everyone goes with now, but anyway, even with the seamless MPLS and with segment routing and EVPN, you would provide layer two circuits for your customers."

This was qualified further …

"If it's seamless MPLS, you're using layer 2 VPN technologies like virtual private wire services, or VPLS where you would emulate an E-LAN and with EVPN, it's the VPN that does all of these."

… and followed by another valuable insight:

"But yes, you definitely provide layer two access to your customer. So you could sell layer two circuits to your customer and ... from what I know from industry, it's more frequent for Tier 1, let's say, providers to provide layer two circuits for other providers – a circuit between countries or something like that. So that's where it's more popular."

Key takeaway: EVPN, signalled with the support of Segment Routing in the control plane, is the successor that takes up the mantle of adopting MPLS in the data plane.

On Layer 0, 1 metro-aggregation technologies

The interview proceeded towards discussion of the results of asking CSPs about their motivation for migrations towards transport systems with integrated DWDM pluggable (away from separate transponders and muxponders), and towards open OLS. I first dwelt on the charts' interpretation, notably that each motive's importance can best be read by reading its skewness, with negatively skewed distribution indicating high relevance and positively skewed distribution indicating low relevance.

"So, first of all, in Cisco, if you want to dig deeper in Cisco, they call it routed optical networks.

So maybe you're going to get some more insights of how Cisco does it because in the meantime I'm working for the system integrator, which is a Cisco partner.

So ... I'm looking at everything through the Cisco lenses.

So, your first of the first big advantage of open XR and everything is the one that DWDM optics can now be packed into switching and routing infrastructure, faceplate and so on and so on.

So yeah, the first big advantage is that these new DWDM optics can now be packed into the routing infrastructure faceplate.

But I want to tell you that this is quite a new subject, let's say. And this is new technology, right?

I don't think you would see it too much deployed yet because if you want to do it, you would need new hardware.

But of course, me as a vendor, I'm always pushing new technology because I want to sell new hardware.

It's good. But right now, maybe existing networks don't need it yet because they still have all those transponders and all that transport infrastructure that may not be out of support and so on and so on.

I think the we will move from the legacy transport networks to the routed optical ones.

***Once those transport devices will the end of life, or maybe if have some greenfield deployment and you're starting from scratch and you're deploying these**."*

Unsolicited, Ovidiu then proffered reduction in OPEX, arising out of lower electricity costs.

"So that's one thing. But the other thing that I would say to a customer to convince him to migrate to these is that you're spending a lot of a lot on electricity because you have devices that are plugged in.

And I think that's a really quite big advantage, especially now because you're spending a lot on OpEx, *especially now when the electricity bills are higher and higher and electricity is.*

Quite expensive. So that's also an advantage.

It's not just that you can plug it in your router, but it's also that you're moving away from the transponder and [you're saving money on the power which you would have, which you would have consumed on the transponder]."

Ovidiu proffered more detail, which was gladly welcomed.

Key takeaways:

1. DWDM is new technology, unlikely to be widely deployed because it requires new hardware.
    a. This helps to explain the dominance of routed optical networks.
2. Apart from operational simplicity, another significant advantage is reduction in OPEX, from saving electricity on separate transponder equipment.

On technology stacks for metro-aggregation

It was time to address stacks of technology and I took the opportunity to ask about the interpretation of "routed optical networks".

> "*I don't think they thought about Cisco's routed optical networks when they chose this.*"

This, of course, corresponded to Eduard Vasylenko's claim about the lack of association between the term "routed optical networks" and Cisco's use thereof. This left me with an attempt to understand what might have been an interpretation, so I asked for clarification, hoping it might correspond to some colloquialism among the community of CSPs. This would allow me to pin operators' choice to the meaning behind the colloquialism. At first, difficulty was found, as may be read from the extract below.

"It's a hard question. So what did they think when they chose this optical networks over Ethernet.

So I think of without ROADMs Let me have a look at the others.

So, DWDM with SDH Ethernet and IP/MPLS – this is quite traditional for big service providers.

…

I don't know why they chose it, why so many of them chose routed optical network."

I observed that "routed optical networks" was the only choice, excluding the legacy SDH/SONET case, that excluded ROADMs. That drew the next comment.

"And also this question is about metro aggregation, so how you're aggregating them again.

You might not need DWDM if you have a small metro network, right?

And maybe that's what they thought about, no ROADMs.

So. Yeah. No wavelength division multiplexing or maybe something like that.

Just some fibers that you deploy between on small distances, between some routers and maybe that's what they were thinking they had in their network."

Key takeaways:

1. Routed optical networks are most likely interpreted as fibre drawn between routers.
2. It was reaffirmed that WDM may not need to be considered, given a glut of fibre.
3. SDH/SONET is legacy technology.

Haider Khalid

On clarity of the questions in the questionnaire

Overall

As regards the questions, I asked "*whether you feel they were clear, whether they were open to interpretation, and whether there are things that really needed to be improved*".

Haider's difficulty was limited to those questions regarding "transmission", which is out of his scope.

"*Most of the questions were clear ... some of the questions that I've found a bit of ambiguity in them not because the question was ambiguous ... I think it was because probably I was not too familiar with those technologies, like the question related to the WDM type technologies ... probably because I'm not familiar with the technologies ' So I won't blame the question. Maybe the question is clear.*"

I then proceeded to enquire about the dissemination of knowledge about standardized reference points (T, U, U1, V, A10) among network engineers. While reference to these RPs was always accompanied by an explanatory graphic in the questionnaire, I wanted to understand obstacles to apprehension both for the survey's purpose and to form my own understanding of jargon familiar to the groups within my research's scope. Haider answered as follows.

"So the diagram itself is very, very clear. So I can see the IP/MPLS based aggregation networks and the way you have drawn these cloud networks, this shows the IP/MPLS backbone aggregation. I think the terminology that you have used like U1 and V, I think these are the terminology that we don't use in industry standards.

So what we use is the UNI and the NNI.

So the UNI is the user-network interface and the NNI is the network-network interface.

So I'm not sure if these two are related to the same that I'm talking about U1 and V1 that you are saying.

Then the traffic flow I'm not able to understand from this diagram because you haven't shown any arrows for the traffic flow, if this is related to some traffic flow from the user side."

These comments cast some doubt on the use of reference points among CSPs' technical personnel, and so I pursued this further, and referred to another of the graphics.

"The reason why I used the T and the U and the V when they are standardized by the ITU and the broadband forum ... if they are used properly, they tend to be specific with regard to the point in the traffic flow which they are referring to ... [moreover], as regards the graphics, well, there are a couple more like that."

Haider's response affirmed the value of intrinsic clarity in graphics.

"I think over here, the good thing is that you have mentioned the terminologies, like you mentioned the device at the right hand side, like the PC and–the STBs - the set-top box. And you have mentioned at the left –and side - the NSP or the PSTN. So that plays the flow of traffic like it's coming from the user to the network or from the network to the user. So I think this explains itself the terminology that you have used here. That's why I think it's okay."

Wrapping up, I asked:

"[M]aybe we can close this part of the discussion by concluding that, overall, you feel that the structure and the text of the questions is at least not evidently problematic. Is that a fair conclusion?"

Haider concurred.

"I agree. I agree. I think that's all right. Yeah."

Indeed, while discussing aggregation results, I referred to the V reference point and the A10 reference point – to which Haider replied:

"from the access to the provider edge".

Haider's recognition was based on his view of the TR-178 graphic (Fig. 72). This was a good indication of the interpretability of the graphic presented as the basis of a number of questionnaire questions.

On the interpretation of the term "routed optical networks"

Haider commented as follows.

"I would say that if somebody who has not worked on Cisco and is from like Juniper or maybe from Ericsson or Huawei or other vendors than Cisco, ... I think that person would not be able to understand this because this is more like into vendor stuff. So I would say that it's better to keep it industry standard

so that it entertains all the audiences like without any difference. So all the audiences should understand what is like being asked."

However, he qualified this with:

"[Y]ou know, Cisco is the actually the go to vendor for everybody. So, yeah, some use Cisco terms quite interchangeably."

Key takeaways:

1. Reference points may not be widely known among CSPs.
2. Routed optical networks is not a term that has precise meaning.

On Layer 2 and higher metro-aggregation technologies

Shortly after introducing the results to the question about layer 2 and higher aggregation from access node to service edge, Haider proffered the following.

"Just to let you know, that provider edge is the term that we use in the ISPs. So, so because I've worked for multiple ISP, then we always have always used and heard provider edge ... service edge is used but not as frequently as provider edge."

A significant contribution followed, when Haider interpreted the results.

"I would agree with the right side [(the SGA results)], because in most of the networks, seamless MPLS is being integrated ***now*** *in many networks. But still, most of the - I would say probably 70 to 80% - of the networks around the world, because I work in different parts of the world and I know in the Middle East, in the Asian region and even in the European region, most of the networks, they do this QinQ and bridging stuff between the access and aggregation. So,* ***the MPLS starts at the provider edge*** *and then it'll go so: from the provider edge down to the customer side, the access and the aggregation, they prefer using QinQ and the provider bridging. Some ’f –he ISP's - they do seamless MP’S, but it's not as common as the other technology. So, I would say 70 to 80%, where I work in the Middle East and in Asian like South Asia and even in the European countries, I've seen this provider bridging and QinQ in most of the access."*

Key takeaways:

1. When Haider’s comments are combined with Ovidiu’s and with the statistics, it emerges that the subset of Regional and Tier 1 operators which are active in the access and aggregation, are more likely to operate:
   a) IEEE 802.1Q-2022 aggregation all the way up to the metro core, and
   b) IP/MPLS switching in the metro core.

On Layer 0, 1 metro-aggregation technologies

I presented the results of the question asking CSPs to rank their motives for migration towards integrated DWDM pluggable and open OLS. Haider agreed that the facility to pack DWDM pluggables densely is an important motivator, as well as the importance of mixing of grey and coloured pluggables. For the NOG respondents, these were the motives that leant most and second most towards high relevance.

"[W]hen I was working in the Middle East, so there was DWDM over there in the Middle East, the ISP that I was working for. And they used DWDM technology ...the thing that the line card can now carry a mix of the grey optics as well as the DWDM optics ... that is also a very important factor, because previously when we had GBIC cards, those were big, big cards. We couldn't insert those cards in the back end switching and routing blades on the routers. So I think this is also a very important factor for DWDM to be implemented in the networks that because of the optics size to be reduced as the same as the grey optics, that is a convincing factor for DWDM to be successful."

I tried to elicit another perspective on the role which OTN would play. Haider's response at first indicated conflation with other optical networks that are deployed downstream of the access node. Nonetheless, his answer was a useful reaffirmation of his previous elaboration on the co-distribution of technologies by metro area segment.

"So, I would say that current ISP, the next generation networks, they are not going towards OTN because they are moving towards XGS-PON. So, like even in the in the European countries like in the UK that I'm working the company currently I'm working for. So that is one of the biggest ISP in the UK. So, they are also moving towards XGS-PON now. So, they are like, you know, in XGS-PON we have the OLTs, the optical line terminals and that is the fibre to the home. So, we provide fibre services from the service end down to the customer premises. So, it's a direct fibre connection from the core to the customer edge. So that technology, the technology that we used to provide that previously it was GPON, but that was 1G downlink and I think 1G uplink. But now with the advancement in technologies, we are shifting towards XGS-PON. So that is symmetric 10G both ways from the customer edge to the user side and from the user edge back to the core. So, for that technology to be implemented, we need OLTs and ONTs. ONTs are installed at the customer end - optical network terminals and the optical line terminals, they are installed in exchanges at the access as an access node and from all the different customers they have, each customer will have its own ONT where the fibre terminates directly in ' customer's home and from that, ONT, that fibre goes to the exchange to the OLT and one OLT can serve one area. For example, if in Manchester we have an exchange, the Manchester exchange will have an OLT that will be serving thousands of customers in Manchester, and from that OLT then we are facing towards the network side. From that OLT, we get the traffic to the aggregation layer, or you can call that the Metro network, and then from there we can get the traffic to the provider edge in the core. Once

we reach the provider edge, then we have MPLS and we switch traffic on the basis of MPLS. So, this is what we are moving to now."

Summarizing:

- XGS-PON from customer edge to AN ("exchange"),
- aggregation (presumably 802.1Q S-VID (service VLAN identifier) and C-VID (customer VLAN identifier)),
- provider edge to metro core MPLS label-switched paths.

For example, in such a network, an L3VPN can be supported through marking customer subnets by a particular {S-VID, C-VID} combination.

To try and obtain more insight on the role of OTN as an aggregation technology, I described it as a successor (in the sense that it employs TDM) to SONET/SDH. This elicited recognition.

"Yes. Okay. All right. It is being used. I agree. Yeah. Yeah ... I would say that in future this will be dismantled. That's my understanding. Yeah. Yeah. It seems to be that way ... So yeah. Yep. So I think, I suppose, they're moving towards packet based now. So they're moving towards packet based."

Later in the discussion, Haider indicated that SDH is still in use in his organization, so I asked:

"Is there any thought of dropping it in favour of something else?"

The reply was unequivocal …

"Yes, they are. They are planning to replace that with DWDM all across the transmission."

… and, moreover, the reply was coherent. The organization involved is a large one, with several million subscribers. The need for return on investment in SONET/SDH technology would have supported its retention well past some form of collective realization in the sector of CSPs that the technology had been superseded.

Topology was the next point of discussion; I sought Haider's perspective on the results about the question whether meshes are likely to be more common in the future among metro-core nodes.

*"I would say it entirely depends upon the design of the network. So who is the designer of the network? So I would say that the ISPs that I have worked with, **so probably 80%, 70 to 80% were using full mesh in the metro network**, but there was some ISP's they were not using, they were using some other topologies like ring topology in the metro. But again, because it depends upon the design of the network, like how big is the neighbourhood? If a network is too big, then full mesh is not recommended because ... if you're doing full mesh in a very big network, then it means you have to have lots of connection' and that's not scalable. So, if you have a smaller ISP, full mesh is always definitely recommended, but*

even with bigger ISPs, I've seen full mesh. But, I think, for that, to implement it requires a lot of resource, a lot of effort to get full mesh cabling in large networks. But again, full mesh is something that is being used a lot in that area."

On this reading, even access nodes aggregators (e.g., the Ethernet switches upstream of the V RP) are fully meshed in 70 – 80% of CSPs. This does agree with one anecdote which I can personally relate with regard to a local CSP. Rings, therefore, while apparently convenient, do not give CSPs the desired level of assurance on service availability.

Key takeaways:

1. While rings are commonly used as examples in literature (see, for example [108, Ch. 17], [342]), mesh interconnection of access nodes aggregators (for emphasis's sake: access nodes are devices like DSLAMs, CMTSs and OLTs, which aggregate subscriber lines) is at least equally likely.
2. Another example of SONET/SDH's removal from aggregation technology stacks, was given.

On technology stacks for metro-aggregation

When asked about technology stacks for metro aggregation, Haider replied as follows.

"So, I'll tell you, because I worked for one of the biggest ISPs in the Middle East and I'm working for one of the biggest ISPs in the UK. So, we have a subscriber base of around 6 million subscribers. So, this is a big ISP and both ISPs in the Middle East and in England, both are using ... IP/MPLS plus Ethernet plus DWDM ... this is what being used for both big ISP that I worked for in England and in the Middle East. So, so I would say that this is the more preferred."

I asked Haider about use of ROADMs; while he was unable to reply, his answer drew attention to the classical division between "transmission" people and those (like Haider) who work at higher OSI layers (MPLS and IP, notably). Now, other than the implementation of DWDM for higher utilization of point-to-point physical cabling, all implementations of DWDM requires some form of filtering to select specific wavelengths. While there are some variants in the genre of optical cross-connects (OXCs), the ROADM, notably in its colourless-directionless-contentionless-flexible grid form, represents the state of the art of the genre. Haider's reply – in so far as it seems to correlate CSP size to technology stack – therefore reaffirms the correlation between answers given to this question by SGA respondents, and large subscriber base size.

Key takeaways:

1. Regional and Tier 1 CSPs active in the metro area are adopting IP/MPLS over Ethernet over an optical network comprising DWDM links that are optically switched using ROADMs.

Dave Eilert

On Layer 2 and higher metro-aggregation technologies

Dave’s position with AT&T gives good visibility into the current implementation of a large CSP occupying both regional and Tier 1 roles.

“So generally, what I see is from the edge down to, let's say, the access site, it's most mostly like Q-in-Q, it's mostly Ethernet over VLAN with some kind of VLAN ... Then, that connects up with MPLS in the backbone. So, it'll be like MPLS in a backbone, and then it'll be like some kind of Q-in-Q down to the site.”

Key takeaway:

1. Further emphasis is made on what was observed earlier about aggregation in regional operators’ networks:
   a. IEEE 802.1Q-2022 aggregation all the way up to the metro core, and
   b. IP/MPLS switching in the metro core.

On Layer 0, 1 metro-aggregation technologies

I addressed the motives for migrating towards integrated DWDM pluggables.

“The blue bar is huge, right? Because if you can plug it right into, you know, your access nod’, now you're eliminating a piece of hardware that's a potential point of failure, right? So, you know, a lot of times now the fibre will come in to a facility, and you have to have some kind of fibre box or some kind of piece of equipment that is like a fibre converter, right? You plug it into a fibre port on a switch, and then it comes out the other side of the switch as copper, and then the copper plugs into your router. Right? So, if you could plug the fibre directly into your router faceplate, I mean that's huge, right? You're eliminating that point of failure.”

I asked Dave what he thought about OTN being displaced by packet networks.

“You're talking about something kind of TDM wise ... Oh, yeah, yeah. I agree 100%. Right now, what happens is, you know it's’ it's, it's all Ethernet.”

On technology stacks for metro-aggregation

"So, I find the first four bars, SONET and Ethernet plus MPLS … I find them more in the local PTTs[53]*. They're running, kind of because it's you know, it's a large investment to, you know, overhaul your network. Right? So, they're using what they can for as long as it is as long as they can use it. Right? So, every once in a while, we come across Ethernet over SONET, or Ethernet over ATM. And, you know, so, you know, I find that more in the local providers than us. We don't have anything like that. It's just strictly Ethernet."*

Key takeaways:

1. Return on investment is a key criterion in determination of the rate of penetration of replacement technologies. SONET/SDH is particularly hard to displace in the "local PTTs" – these are the CSPs of a smaller, regional scope, the descendants of what are colloquially referred to as the "Baby Bells", when AT&T was broken into Regional Bell Operating Companies.

8.4.2 Written media: e-mails and reports

The two broad objectives (discussion of the graphical summaries of the results and assessment of the objective clarity of the questionnaire's questions) were further pursued through written media. This approach facilitates reflection on both parties' sides, whilst lacking the immediacy obtained in face-to-face interviews. For convenience's sake, the portion of Table XX pertinent to written media, is reproduced below as Table XXI .

Table XXI	WRITTEN RESOURCES COLLECTED DURING DATA STAGE OF QUALITATIVE SURVEY	
E-mail thread: Mark Tinka[54]	[334]	https://drive.google.com/file/d/1eltC-QaLPiZitrlngEWAUS5gtiimY-jw/view?usp=sharing
E-mail thread: Philip Smith[55]	[323]	https://drive.google.com/open?id=1bN0RS_M93UojmdA0oYMYqJI9Ql_uM8Hk&usp=drive_fs
E-mail thread: Daniel King[56]	[332]	https://drive.google.com/open?id=1ayHhTdpfdIWMZuyLPTapKDZS1-cUFIRM&usp=drive_fs
E-mail thread: anonymous	[318]	https://drive.google.com/open?id=1el7oH18k_dWsupVPNjCfGHSAaa_7tkX3&usp=drive_fs
Written assessment: Haider Khalid[57]	[331]	https://drive.google.com/file/d/1gjth5v8W2x4Dll_l5oyPW9n39uT9fnGc/view?usp=sharing

---

[53] Postal, telegraph, and telephone service
[54] https://www.linkedin.com/in/mark-tinka-5b03055/
[55] https://www.linkedin.com/in/philip-smith-154502/
[56] https://www.linkedin.com/in/danielking/
[57] https://www.linkedin.com/in/haider-khalid/

Mark Tinka

Mark's support in crafting the questionnaire precluded any discussion on clarity of the questions. The results were within scope of discussion; a first reading drew some scepticism, due to the dominance of ADSL2+ in SGA's sample, both as the largest and fastest-growing access technology. The ensuing exchange with SGA's representatives accentuated the importance of balancing quantitative surveys with qualitative surveys. Therefrom, it emerged that with ***regional and Tier 1 CSPs*** dominating SGA's, and with further suggestion that these were indeed incumbents in their markets, the SGA sample includes CSPs who are still reaping returns from their investment. One particular extract of the thread is particularly pointed.

"So, I've read their–response - I suppose it makes sense, because AT&T and Verizon, especially, are legacy operators with tons of copper in the ground that they are, most likely, still monetizing.

While they are likely to have all of that legacy infrastructure that comprises a huge part of their inventory, it does not necessarily mean that the world is not moving on to fibre, Ethernet and DWDM.

*I think SGA would need to consider, for the future, how to obtain data from a wider set of operators (of varying sizes and scope), because a lot of the next-generation deployments don't generally tend to happen (fast enough) at the legacy incumbents. That, I think, is why the data is quite different from the *NOG responses, because the *NOG responses cover a substantially wider base of operators, most of whom are not normally representing legacy incumbents."*

This observation is reinforced by Dave Eilert's, who, in the context of reference to SONET/SDH, had reminded me about the need to exploit investment in infrastructure. Indeed, SONET/SDH is still in use by 10% and 24% of NOG and SGA sample respondents, respectively.

Key takeaways:

1. Return on investment is a key criterion in determination of the rate of penetration of replacement technologies. The observation here arose in the context of an access technology (copper-based ADSL2+).
2. The SGA sample's response is representative of incumbents with legacy infrastructure still being monetized.

Philip Smith

While Philip drew attention, as the anonymous respondent did, to uncertainty with the representativeness of the results, he asserted that the results did not jar with his understanding of current metro area networks and developments thereof. Moreover, in the process of evaluating the credibility of the data, a fresh perspective on the two groups was offered: NOG respondents are likely to include those with a higher degree of autonomy in taking decisions than those from the large incumbents. The salient extract is reproduced below.

"Yes, SGA has given you a guarantee of the biggest providers, probably what we'd call the national telecoms (in the old days). My feeling from the industry is that those folks will have much less scope to do what "they want to do", in that technical decisions about network deployments will be made by their vendors. I remember similar in my Cisco days – the biggest deployments by big operators (at least for access) were by and large handed over to the vendors to make a proposal and then implement ... So even if the CTO of a "Tier-1" responded, they'd have not a lot of say in what is made available or what the future strategy might be, apart from choosing the options their vendors present to them.

Doing the survey by open request to the NOG community will mean you'll get the smaller operators and the operators who are not driven by their vendors - the big private operators for example. Is it representative? Well, who knows, but we are stuck with those who are willing to volunteer their time.' Again, I've helped these providers over the years, and they are much more determined to do what is right for the customer, the best and most reliable implementation, most cost effective for them to implement and operate'...

So, I'd summarise that you have two valid data sources in your survey here, but from parts of the community that have different/opposite outlooks on how they deploy infrastructure. Both are valid. TBH I'm more comfortable with access infrastructure that has been designed by the provider than one that is delivered by t–e vendor - my bias after years and years of working helping build Internet service provider networks ... Thanks for sharing–all this - it was an interesting read. I can't think of anything that would invalidate (or cause question) on either the NOG or SGA survey. They are going to be as representative as you can get with a voluntary response."

I also invoked Philip's support on the issue of interpretation of the problematic term "routed optical networks". In view of the significance of the technology stack's impact, I have deferred citing his contribution to resolution of the issue, to the analysis.

Key takeaways:

1. Given the observed difference between the two samples, the results match expectations.

Daniel King

Limitations and ambiguities in the questionnaire were addressed directly, as follows.

"Nothing obvious. I suppose a few questions on planned network evolution might have been interesting; for instance… Juniper is proposing Seamless Segment Routing; they are keen to continue the end-to-end inter-domain/AS principle of seamless MPLS with lightweight traffic engineering features of SR. I know at least two mobile operators working with Juniper on this architecture – although several standards gaps exist, such as BGP-CT maturity."

Indeed, Daniel's observation is correct; however, since the scope of research was the data plane, and the questionnaire required 10 – 12 minutes [324] of a respondent's time, I felt the need to avoid digressing.

I also asked how well the results match his perception of trends.

"The service distribution of the existing customer base and current/predicted service growth meets my expectation. Although, I have not discussed specific customer service demand with an operator for several years. However, the results match the recent trend of standardisation activity, including discussion in the ETSI Fifth Generation Fixed Network (F5G)."

Given Daniel's background as witnessed by his participation in several RFCs[58], I asked for his opinion on the widespread choice of seamless MPLS as a layer 2+ aggregation technology.

"Service evolution for residential customers has recently transitioned from a "Broadband Era" (XG PON), to "Ultra Fast Broadband" (NG-PON), and we are now in a "Gigaband Era" (50G EPON). It is worth noting that the operators not only wanted to increase the number of users supported and provided more bandwidth, but an operator must also be able to dedicate bandwidth for 5G fronthaul and provide backward compatibility with EPON/10G-EPON and GPON/XG PON, whilst reducing operation costs (especially reducing overall system power consumption and cooling costs).

Why is seamless MPLS so popular with your respondents? Transceivers now support a higher power budget, reach, flexible grid transmission, and system power cost. Still, they can also be used in multi-layer packet routers that already have MPLS fast-path forwarding ASIC/FPGA's. In addition, these GPON/EPON transceivers provide a server layer for seamless MPLS architectures – which can be managed using a single control plane to simplify operations and red–ce costs - supporting ultra-fast end-to-end Internet services across a range of residential, business and vertical user and application requirements (BW, protection and latency).

Importantly, to reduce operational costs, setting up fine-grained services across multiple domains (end-to-end) whilst collapsing control plane architecture is a huge advantage. Ultimately, I think energy efficiency constitutes an increasingly significant challenge for network operators; not only does

---

[58] https://datatracker.ietf.org/person/d.king@lancaster.ac.uk

it directly reduce the operational expenditures of operators, but it also lowers both carbon emissions and environmental impact."

Key takeaways:

1. Multi-layer packet routers are facilitating delivery of differentiated end-to-end services – and these services' availability benefits from the smaller domains facilitated by seamless MPLS.
2. The results match expectations.
3. The questions are clear.

Haider Khalid – report

In addition to the face-to-face interview, Haider accepted to dwell further on the content and wrote a brief report [331]. The key takeaways from the report are reproduced below.

Key takeaways:

1. UNI and NNI are easily recognizable terms; U1 and A10 are not.
2. In similar vein: provider edge (PE) is recognizable as the edge of access. This calls to mind the Stage 2 segmentation model.
3. The results match expectations. Haider cited the following as particular cases of the agreement between results and expectations:
    a. distribution of deployment of access technologies;
    b. Ethernet as a "major layer 2 backhaul";
    c. MPLS as dominant switching technology, especially when core is taken into account, and
    d. location of video BNGs as close as possible to the end-user, to save bandwidth.
4. No ambiguities in the questions were detected (though the graphs presented to Haider for his analysis, were found lacking in clarity).
5. Limitations observed concerned the desire to extend into questions on SDN. Here, my defence is the same as that offered with regard to Daniel King's observation on scope of the quantitative survey.

8.5 Analysis

This section is divided into two parts. The first sub-section carries saliencies of the results, obtained as a product of the quantitative data and discussions thereon. Structured discussions were held with named reviewers, while other field experts contributed to specific issues. The second sub-section presents a first set of scenarios that emerge as candidates for full implementational modelling. The analysis closes this chapter.

8.5.1 Saliencies emergent from quantitative and qualitative survey

1) The most common technology stack: routed optical networks

What started as an issue in the crafting of the questionnaire developed into an exquisitely satisfying corroboration, as a product of the technique of qualitative survey. The first hint at the interpretation of this popular choice of metro-aggregation technology stack which CSPs are deploying emerged from discussion with Ovidiu-Mădălin Roșeț [327, N. @38:23]. Clear evidence in favour emerged from Philip Smith [323, N. See mail on June 27th 2023].

> *"Okay, put it this way, folks who are building a network using fibre optics plugged directly into routers know exactly what they are … That's all it is: fibre and more fibre plugged into routers."*

I pressed further and wrote the following.

> *"Cisco's primary purpose seems to be to obviate OTN, but not to dispense with ROADMs. That's an important differentiator with respect to*
>
> > *"fibre and more fibre plugged into routers".*
>
> *I suppose that such implementations, i.e., "fibre and more fibre plugged into routers", can be achieved without ROADMs. I can see that happening with enough dark fibre, or with enough wavelength service purchased from some CSP who ***does*** use ROADMs,"*

Philip's reply solidified the interpretation.

> *"And yes, places where you cannot get enough fibre installed, something else needs to be done. And that's where the solutioneering starts … I'm sceptical about a lot of it – most networks are built with the "keep it simple" principle, and that's not something that flies well with the vendors, who ever they are."*

Re-visitation of an e-mail exchange held earlier [344] affirmed this understanding. I had asked:

> *"Where are operators gravitating towards in their selection between packet- and circuit-transport technologies in the MAN?"*

Furthermore, I qualified the question, as follows (CSP name replaced).

> *"[I]s an optical transport network in use with CSP_Y? I mean, an OTN with various data rates? As opposed to pure packet switching all the way …"*

The answer given was the following.

> *"In Malta given the short distances, up to now it is cheaper for us to deploy optical interfaces directly on fibre, rather than to introduce an OTN/WDM layer. This would not be the case in larger countries where distances are longer."*

This is in good agreement with Arelion's representative's answer, quoted earlier, i.e.:

> *"We investigated every route we have between routers and found out that about a third of the routes are < 40 km & < 400G so that would mean we could replace every metro DWDM we have there with this new pluggable technology and shoot from router to router."*

Therefore, it is reasonable to conclude that, unless qualified otherwise, the term "routed optical network" evokes the simplest operational architecture that supports packet switching, namely:

1. Layer 0: dark fibre
2. Layer 1: a pluggable transceiver, possibly using grey optics such as 10GBASE-ER, or possibly using coloured pluggables in anticipation of their future use.

Given the extensive adoption of Ethernet and MPLS, and the ubiquity of IP, it seems fair to extrapolate further to cover layers 2 – 3 with these technologies. One additional consideration is worthwhile. While discussing an related technical matter with Mark Tinka, my attention was drawn to the extent to which some CSPs take simplicity [345].

> *"[T]here remains a number of operators who run MPLS-free backbones, and are religious about simplicity, and forwarding all of their traffic via plain old IP."*

Indeed, one CSP, while answering the question about the technology stack, opted to answer with his own text (using the "other" catchall), and wrote "dark fibre + Ethernet + IP".

Regardless of whether MPLS is employed to forward traffic in the data plane or not, the convergence of the variety of sources strongly indicates the intended composition of the stack of layers implied by the common choice "routed optical networks".

2) Layer 2 and higher aggregation from the access node to the provider/service edge

When all data – both from quantitative and qualitative survey – are processed, they cohere well and conclusions can be drawn on the state of current networking and the expectation for next-generation networking in these layers of the technology stack. Seamless MPLS was identified as an intermediate step and of current interest to CSPs. It is strongly preferred by NOG respondents as an aggregation technology, but Provider Bridging is most common among Tier 1 and regionals (SGA). Given the larger average subscriber base size of the Tier 1 and regionals, it matches intuition to find that the rate of adoption of newer technology is greater among the group of CSPs (the NOG set) with smaller average subscriber base size. Therefore, on the basis of the data collected from the two surveys, it can be observed that current layer 2 and higher aggregation, in descending order of deployed instances is the following.

1. Most deployed
    a. V to A10: IEEE 802.1Q-2022 Provider Bridging, or QinQ
    b. Beyond A10: MPLS
2. 2nd most deployed: Seamless MPLS, from V to beyond A10

Moreover, deployment of next-generation aggregation will be ordered as follows:

1. Most deployed: Seamless MPLS, from V to beyond A10
2. 2nd most deployed
    a. V to A10: IEEE 802.1Q-2022 Provider Bridging, or QinQ
    b. Beyond A10: MPLS
3. 3rd most deployed: Segment Routing controlled multi-domain MPLS

It is difficult to anticipate that SR-controlled multi-domain MPLS could easily overtake both other aggregation sets. This is due to the paradigmatic change represented by SDN, which requires cultural change within CSPs. Any further comment at this point would be little better than pure conjecture.

3) Distributed Access Architecture – a split along sample boundaries

By far the largest percentage (44.4%) of NOG respondents are not planning a distributed access architecture; in contrast, it is the smallest percentage by far (2.1%) of SGA respondents that are not planning a DAA. One respondent in the NOG sample helpfully opined that "[a] distributed access architecture is applicable in large metro areas. In small metro areas (in terms of geographical coverage) it is typical to concentrate equipment in a few PoPs.". If the smaller mean, median and mode of the NOG sample are indicative of the geographical coverage of CSPs in the sample, then the difference between the two samples might be explained on this ground.

The importance of DAA lies in its inherent infrastructural demand for a new RPI-N: that between the two actors identified in [214, p. 67]. Even if the Virtual Network Operator and the Infrastructure Provider are "subsumed into a single entity", a RP[59] must be identified for the interconnection within the designated real estate: the "NFVI-PoPs such as central offices, outside plant, specialized pods" [214, p. 16].

---

[59] A RPI-N is a specialized RP. If roles on either side of the RP are subsumed by one entity, then the scope for an RPI-N is replaced by that of an RP.

4) Correlation between DAA and virtualization; between DAA and customer-proximal MEC

65% of NOG respondents do not plan to fully virtualize; the same percentage does not plan to employ DAA for the majority of HHP. Moreover, 44.4% of NOG respondents have no plans at all for DAA. On the other hand: all SGA respondents plan to fully virtualize within 5 years or less, or have done so already, and the same percentage – 2% – does not plan to use DAA as that which does not plan to serve the majority of HHP with DAA. A similar correlation exists between DAA and Option 0 (see Fig. 72) MEC nodes. In more detail than is visible in the charts: of the 32 NOG respondents who are not planning any DAA, only 5 indicated that they are deploying MEC in locations close to the customer (option 0).

These two correlations seems to justify the observation carried in [214, p. 67], that “virtualising broadband access nodes can be exploited by the co-location of wireless access nodes in a common NFV platform framework … thereby improving the deployment economics and reducing the overall energy consumption in the combined solution.” Admittedly: a wireless access node is not a DAA node, but it must be conceded that any extant wireless access node is a candidate for extension of its role into that of DAA node.

8.5.2 Scenarios

The number of aggregation and core scenarios is too large to handle well all at once. Rather, it seems more useful to use the common implementations as the bases upon which to support future modelling work. Of course, the surveying techniques used here are only part of the methodology (see the framework section), but they are a significant and valid part. In this sub-section, the results of surveying will be used to compile a set of aggregation and core scenarios for modelling.

1) Layer 2 and higher aggregation

A first scenario: PB, with G.8032 and TR-101 N:1 forwarding

With the larger subscriber-base CSPs (SGA sample), an initial reading of Fig. 78 shows a dominant Provider Bridging (Q-in-Q) component. Some limitations must be addressed before this common choice can be interpreted into a realizable implementation.

Limitation: number of service deployments

Used without MPLS, PB is limited to roughly 4000 different service deployments. Indeed, BBF TR-101 Issue 2 [149, p. 32], which describes Ethernet-based broadband aggregation states explicitly that “Aggregation Switches will only forward based on S-Tags” and requires the “uniqueness of a S-Tag across ANs”. For example, if an enterprise customer were to purchase an attachment circuit (AC) connecting the customer edge (CE) switch (the UNI-C device) to the provider's UNI-N device in a local exchange, then that customer would be assigned one of the 4092[60] assignable VIDs.

Resolution: N:1 forwarding model

The limitation on the number of S-TAGs is most significant if the customer must be isolated from other customers with his/her own S-TAG; with the N:1 (N ports, 1 VLAN) forwarding model, the low revenue residential customer can be aggregated with far less infrastructural (in terms of S-VIDs used) cost. Therefore, PB that backhauls traffic back to the Internet BNG is not subject to the limitation of 4092 different customers.

Limitation: time to reconfigure topology in case of link failure (convergence time)

A serious concern in aggregating using purely Bridged Networks (Ethernet in layers 1 and 2) regards topology reconfigurations. Topology reconfiguration is required whenever a (layer 2) link fails, whether because of port, media or channel failure. IEEE Spanning Tree Protocol (STP) is well known to have unacceptably long convergence times.

Resolution: simple guarantees on restoration

There are two principal technology families that support fast topology reconfiguration:

1. STP's successors (Rapid Spanning Tree Protocol (RSTP), Multiple Spanning Tree Protocol (MSTP) and Shortest Path Bridging (SPB))[346], and
2. the ITU-T's G.8032 Ethernet ring protection switching (ERPS) [347] and G.8031 (Ethernet linear protection switching) [348].

G.8032 ERPS states simple guarantees on convergence (topology reconfiguration) that are familiar to CSPs: sub-50 ms under specific load conditions and a limit to the number of nodes. There is no similar, evident guarantee with the IEEE standards, although RSTP and MSTP have significantly reduced the time required to adapt active (forwarding) topology to match port state. In the context of the qualitative survey, I have attested to CSPs' perception of longer convergence time with STP and its successors. Academic publications tend to follow suit in this perception of guarantee on availability [349], [350]. I must add that there appears to be no ***empirical*** evidence of superiority of ERPS over RSTP, e.g., by comparison of convergence under identical topologies.

Therefore, despite its limitation to ring topology, ITU-T G.8032 provides a topology reconfiguration technique that complements IEEE PB and, furthermore, BBF TR101 Issue 2 VLAN N:1 forwarding model mitigates IEEE PB customer discrimination limit.

---

[60] The VID field is 12 bits wide and VIDs 0, 1, 2 and FFF are reserved.

Canonical support for the scenario

Ethernet aggregation is defined in TR-101 Issue 2[149], wherein the access node is S-VLAN aware and ***must*** support adding both S-Tag and C-Tag, as necessary. Fig. 97 is a reproduction of [149, Fig. 3]. It serves to support understanding of how PB Q-in-Q can be deployed on the access network: since the access node is S-VLAN aware, and there is no other aggregation technology in use, then upstream traffic is PB Q-in-Q. As stated earlier in this sub-section, S-Tagging of residential customer traffic is likely to follow the N:1 model, where several residential subscribers' ports on the access node are mapped to the same (S-Tagged) VLAN.

Anecdotal support for the scenario

A good example of this kind of deployment – and noteworthy corroboration that it is practiced by large-subscriber-base-size CSPs – may be found in [351], wherein it is claimed that "[f]or very large service providers with millions of subscribers this sort of approach normally works well".

A related technology, with no evident market support

A note must be made about an L2 technology that scales better than Provider Bridging, i.e., Provider Backbone Bridging (PBB, IEEE 802.1Q-2022, [346]). However, in two independent e-mail exchanges which I conducted, PBB was described as "all but abandoned which I'm totally OK with. PBB was just basically re-inventing MPLS using Ethernet MAC instead of an MPLS shim" [352], and "PBB-TE … was 'hot' on conferences some years back but the market interest has been limited. Telia has not to my knowledge used the solutions in practice" [353].

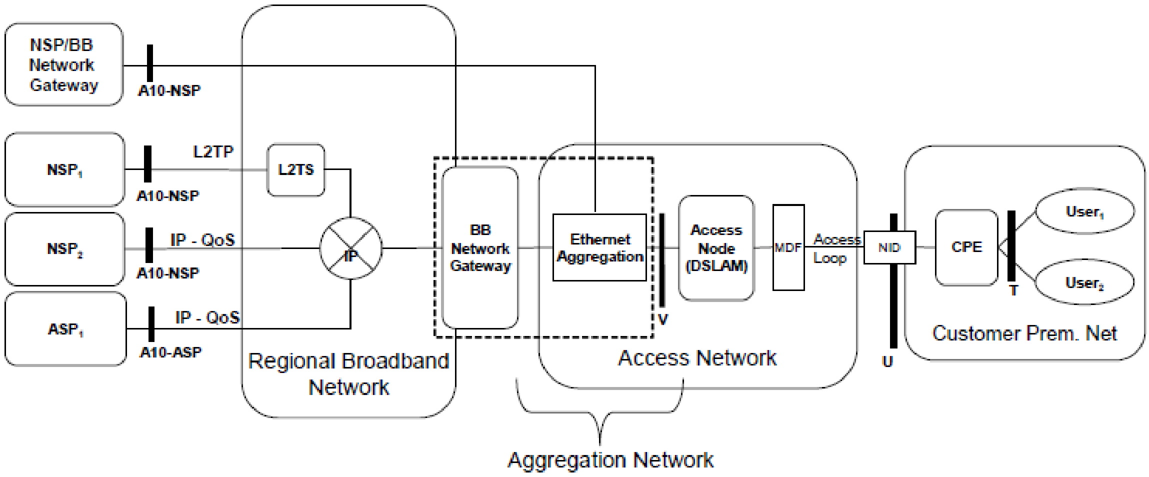


Fig. 97. Simplified schematic of Ethernet-based aggregation of broadband customers

A more nuanced reading, leading to other scenarios

A second look at Fig. 77 and Fig. 78 reveals that, when all the components that aggregate using data-plane MPLS are lumped together, MPLS-based aggregation dominates again. This is less surprising then may appear, as MPLS and PB aggregation are not mutually exclusive. Indeed, the co-existence of MPLS and PB is suggested by the fourth (from top) horizontal bar in Fig. 77 and Fig. 78: "PB Q-in-Q closer to the access with MPLS transport rest of the way back to the service edge". This latter arrangement is one of the "possible combinations of mixing Ethernet and MPLS functions in the various elements, and their corresponding hand-offs" (BBF TR-178 [337, p. 24]). One possible combination will next be presented as a scenario; the combination will then be extrapolated farther to obtain seamless MPLS therefrom.

A second scenario: PB Q-in-Q closer to the access with MPLS transport rest of the way back to the service edge

Now, Fig. 97 shows Ethernet aggregation upstream of the access node, but MPLS aggregation could be employed instead. This is the realm of BBF TR-178, and several observations arising out of that standard's provisions are in order.

1. Fig. 72 (taken from TR-178) shows that an external network – network node interface ***at layer 2*** (E-NNI[61]-L2) may lie anywhere between A10 and V. The means supporting this flexibility is the IP/MPLS-based aggregation (the bottom-most cloud in the diagram). IP/MPLS-based aggregation supports several L2 artefacts, two of which are VPWS and VPLS (see RFC 4664 [355]).
2. BBF TR-178 extends TR-101 Issue 2 through consideration of two new access nodes: an MPLS-enabled access node, and a BNG-embedded access node. These access nodes obviate the need for Ethernet aggregation upstream of the V RP.
3. An aggregation implementation that divides the end-to-end path within a CSP's network, into two or more parts, such that:
    i. one (or more) parts employs provider bridging, and
    ii. the other part(s) employs MPLS

    can be referred to as "segmented" MPLS. Seamless MPLS is the evolutionary step wherein all aggregation within the CSP's network, from access node to access node, or from access node to E-NNI, takes place over MPLS label-switched paths.

As an example of these nuances, a case is now proposed, and illustrated in Fig. 98. The context is the following:

---

[61] E-NNI is used here as defined in MEF 26.2 [354].

1. a single CSP's network's footprint within a single metro area;
2. layer 2 artefacts only are shown;
3. six units of localized real estate, each assigned to a serving area[62];
4. by virtue of its localization, the real estate is a local exchange (LE), or central office, or hub, in the hierarchy of real estate premises;
5. each LE/CO/hub houses:
    a. an Ethernet access node (EAN) for multiplexing of residential subscribers;
    b. zero or more MEF-compliant switches (UNI-Ns) for termination of business subscribers, and
    c. an MPLS PE device;
6. an Internet BNG.

Summarizing, this is a metro area network, and Fig. 98 focuses on the layer 2 architecture of the junction of the edge of the access network, and the edge of the metro-core (which is where the Internet BNG lies). Note that the topology of aggregation, as well as any intermediate real estate, is abstracted for the purpose of clarifying the operation of MPLS.

Each MPLS PE device establishes an MPLS single-segment pseudowire (SS-PW)[63] [329]. Suppose that a business customer attaches to the access node device through an IEEE 802.3 AC, and that the link is a trunk (therefore carrying L2 PDUs with C-VIDs). Furthermore, suppose that the CSP has more than 4092 customers, yet must provide an Ethernet Service Layer. The S-VIDs distinguish the customers at the access node devices, but there are not enough S-VIDs to serve the customer base. The solution here is to use MPLS labels to distinguish the customers between the ingress access node and the egress access node. At the egress access node, the customers can be distinguished using the S-VIDs.

---

[62] There is some redundancy in this phrase, since assignment to a serving area implicitly incurs localization. However, the redundancy is tolerable in so far as it serves to emphasize the geographical scope of the real estate concerned and supports better understanding of the hierarchical arrangement of real estate and network infrastructure.

[63] Note that this is ***not yet*** operation between two MPLS-enabled access nodes. These access nodes are still Ethernet Access Nodes (EANs).

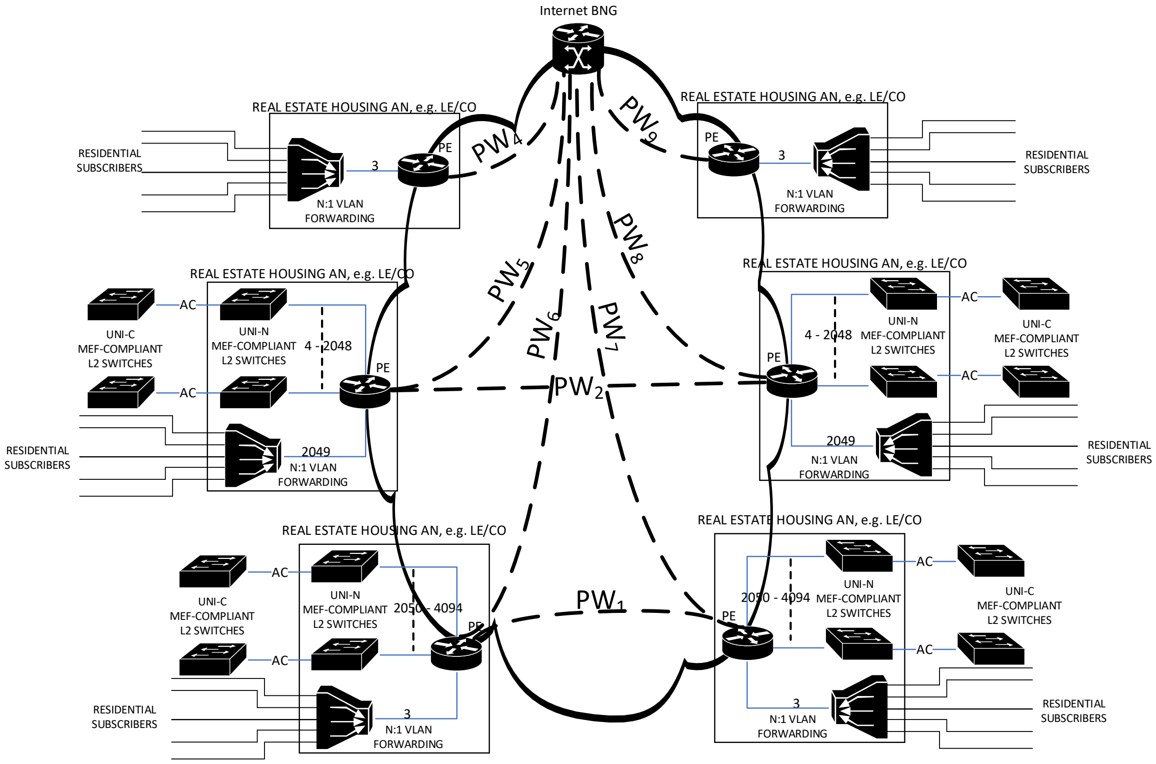


Fig. 98. Supporting more than 4092 ETH service layer customers with an MPLS PE in the CSP's localized real estate.

A third scenario: seamless MPLS

Fig. 77 and Fig. 78 shows that seamless MPLS dominates aggregation in the smaller subscriber-base CSPs (NOG sample), all the way from the access node, into the metro core. That this dominance emerges within the NOG sample is another indication of the agility of smaller-subscriber-base CSPs in the activities of innovation. While seamless MPLS is not radically different in terms of layer 2 architecture, it requires adoption of a newer type of access node: the MPLS-enabled access node. It seems reasonable to expect that the higher capital expenditure and the higher personnel re-training involved would slow the rate at which larger subscriber-base CSPs can adopt seamless MPLS.

Some clarification of the difficulty involved may be obtained from TR-178 (see [337, Fig. 25], reproduced below as Fig. 99), which illustrates the functions embedded in the two types of access node. While the EAN function that interfaces with metro-aggregation at the V RP is the PB functional unit L2F-E [134], the MPLS-enabled AN interfaces at the MPLS adaptation/encapsulation function L2A-M (a function of the LER). The AN is a new device, requiring upgrade, which is better expedited at low unit multiples.

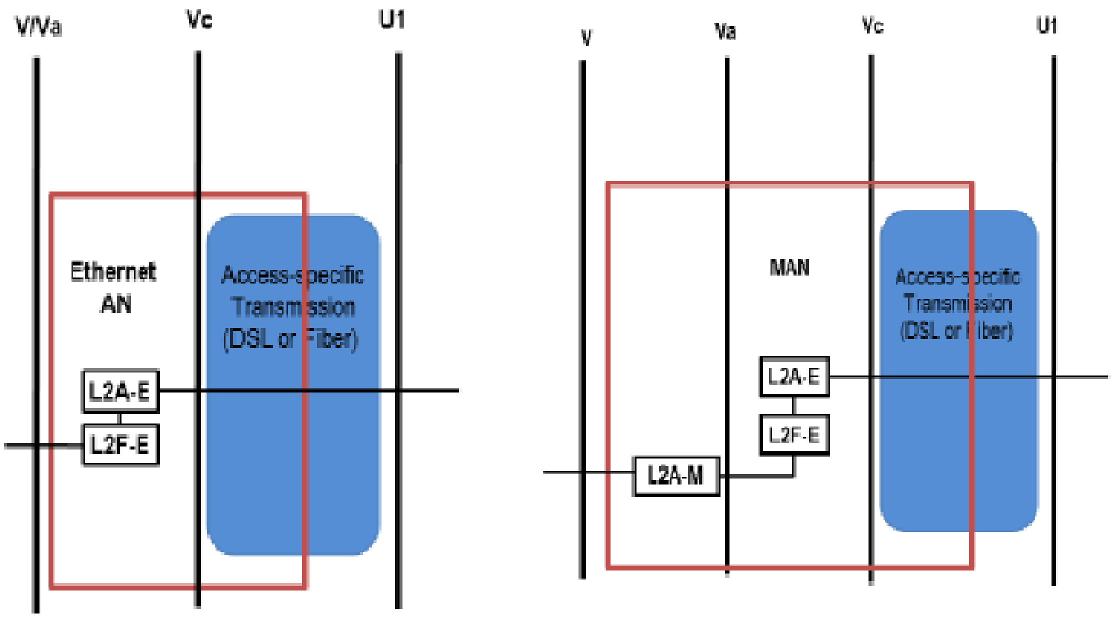


Fig. 99. An illustration facilitating a comparison of the functions embedded in the EAN and the MPLS-enabled AN [337, Fig. 25]

Having established the first scenario as somewhat legacy, yet widely disseminated, and the second scenario as an intermediate step, seamless MPLS has emerged as a desirable next-generation architecture for aggregation. Fig. 100 illustrates the implementation. Note that here, the assignment of S-Tags is provided within the MPLS-enabled access node's L2A-E function, ***at the service of the Carrier Ethernet Network*** (CEN), to support the provision of the EVC artefact for customers.

It should be noted that the abstraction of metro-aggregation is necessary to focus upon layer 2+ artefacts. Now, it is conceivable, and reasonable, that a single metro area may require no more than a single control domain[64], while remaining within limit of convergence time in the event of link failure. Therefore, the PWs shown in both Fig. 98 and Fig. 100 are likely to be single-segment pseudowires (SS-PW), spanning access, metro and core within a single metro area. Notwithstanding the validity of the use case illustrated in Fig. 100, it must be emphasized that the greatest gain in operational simplicity is obtained when an inter-metro PW (or other L2 construct) spanning access-to-access networks, is established with the aid of seamless MPLS.

---

[64] A common and valid example of the scope of a control domain is that within which a set of routers operate in the same IGP (OSPF, IS-IS, say) domain.

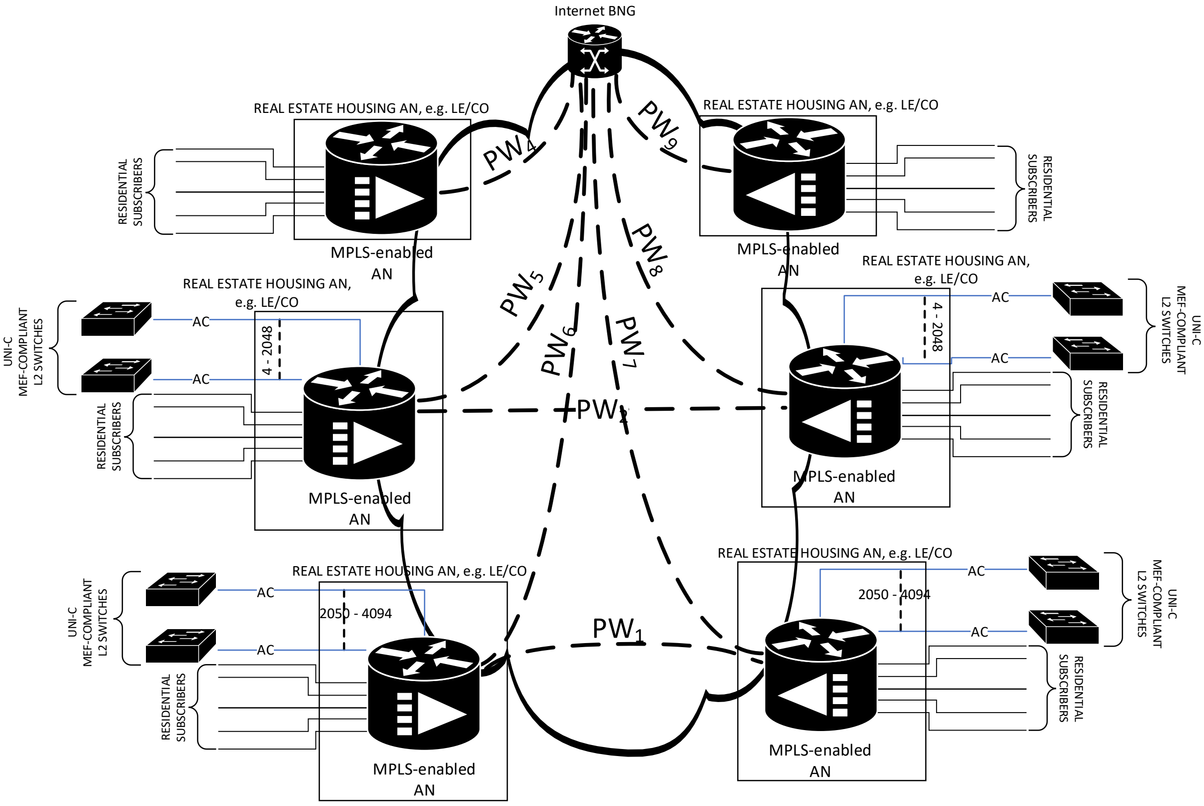


Fig. 100. An implementation of seamless MPLS

2) Aggregation at Layers 0 and 1

Both technology and topology are of interest here; with the surveys' results, the most adopted technologies at layers 0 and 1 can be identified[65]. Technologies thus identified will be applied in conjunction with technologies identified earlier in layers 2+ to establish two scenarios.

Provider Bridging, from V to A10

Table XXIII shows the stack of technologies in an implementation of an Ethernet access node (EAN) in a ring metro-aggregation arrangement of LEs/COs/hubs. The corresponding schematic, suitably overlaid with reference points, is shown in Fig. 101.

The LE/CO/hub includes other devices: a number of other access nodes (optical line terminals) lie just upstream of the optical distribution network's (ODN) S/R RP. Moreover, a mobile network operator (MNO, one type of CSP) has co-located[66] its 5G distributed unit (gNB-DU) and centralized unit (gNB-CU) in the real estate. Both the OLTs and the 5G functional units are uplinked to an Ethernet

---

[65] The surveys convey some knowledge of topology too, but better confidence in the results requires complementing knowledge thus gained with knowledge gained from case studies.

[66] The MNO may be a division of the same organization, or it may be a different organization that has purchased aggregation from the

aggregating device. This is shown separately to accommodate the number of ANs shown in the schematic, but it may well be an integral part of the chassis that includes (the subscriber-facing side of) the access network multiplexer /demultiplexer. Here, this device is the object of attention. Note that Table XXIII shows layers that span the range from optical media (layer 0) to layer 2.

Consider the MNO's operations. Within the 5GS, the NG (logical) RP lies upstream of the gNB-CU and downstream of the 5GC; the entire span of the metro-aggregation network is abstracted. ***As a consequence of*** this abstraction, the protocol stack at the NG-U [356, Fig. 4.3.1.1–1] must be carried unaltered, collapsing the entire metro-aggregation stack into the equivalent of a layer 2 construct (either a link or a MAC Bridge) between the gNB-CU and a UPF in the 5GC. Here, the carrier's aggregation switch linked to the gNB-CU is the MEF UNI-N device. It does not have MPLS capability, but supports Provider Bridging, with stacked VLANs. and the PDUs on the links involved in switching backhauled traffic, are shown in Fig. 102. Note that although a C-VID is shown, the MNO does not need to deliver a tagged Service Frame; an untagged one is supported in the MEF service set.

MPLS backhaul – a case of recursive transport: Ethernet over MPLS over Ethernet

The concept referred to by the terms underlay and overlay is well-known, but the ITU-T's standards on functional architecture of transport networks[67] use more accurate nomenclature in defining (partition-able) layers which may be vertically traversed in a recursive manner[68], down to the physical media layer [161] (or, specifically for optical networks: the optical media layer [194]). Understanding of Ethernet's role within transport is unsound without an understanding of how it can occupy several layers in a stack of transport technologies.

Table XXIII shows the stack of technologies in an implementation of a label edge router (LER) in an access node in a metro-aggregation arrangement. The corresponding schematic, suitably overlaid with reference points, is shown in Fig. 103. As with the previous exposition, the LE/CO/hub includes other devices; the OLTs of various technologies are kept as they were, and again, a mobile network operator (MNO, one type of CSP) has co-located[69] its 5G distributed unit (gNB-DU) and centralized unit (gNB-CU) in the AN's real estate. Once again, the OLTs and the 5G functional units are uplinked to an aggregating device: in this particular case, it is an MPLS-capable device: it is as an LER.

---

[67] Both G.805 [161] – the connection-oriented version – as well as the later G.800 [175] that addresses connectionless as well as connection-oriented transport.

[68] The lowest layer is the transmission media layer network, but this is composed of section layer networks, which include all functions (e.g., transceivers and amplifiers) necessary for information transfer, and the physical media layer network, which is composed of one of the following (a) the fibre, or a channel within the fibre, (b) metallic wire, or a channel therein, and (c) a wireless radio frequency channel.

[69] The MNO may be a division of the same organization, or it may be a different organization that has purchased aggregation from the

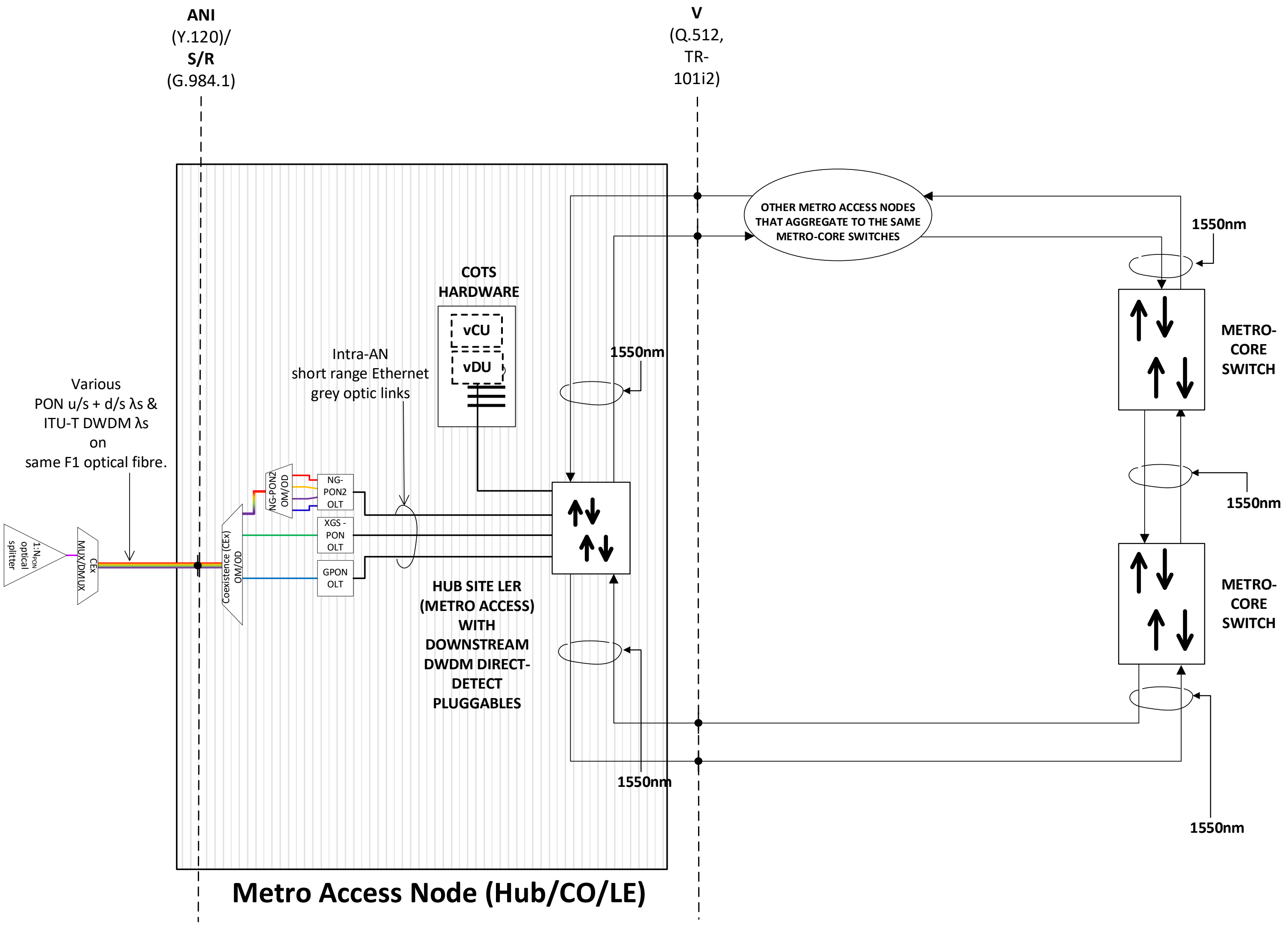


Fig. 101. PB backhaul from access node to metro-core over an Ethernet ring topology

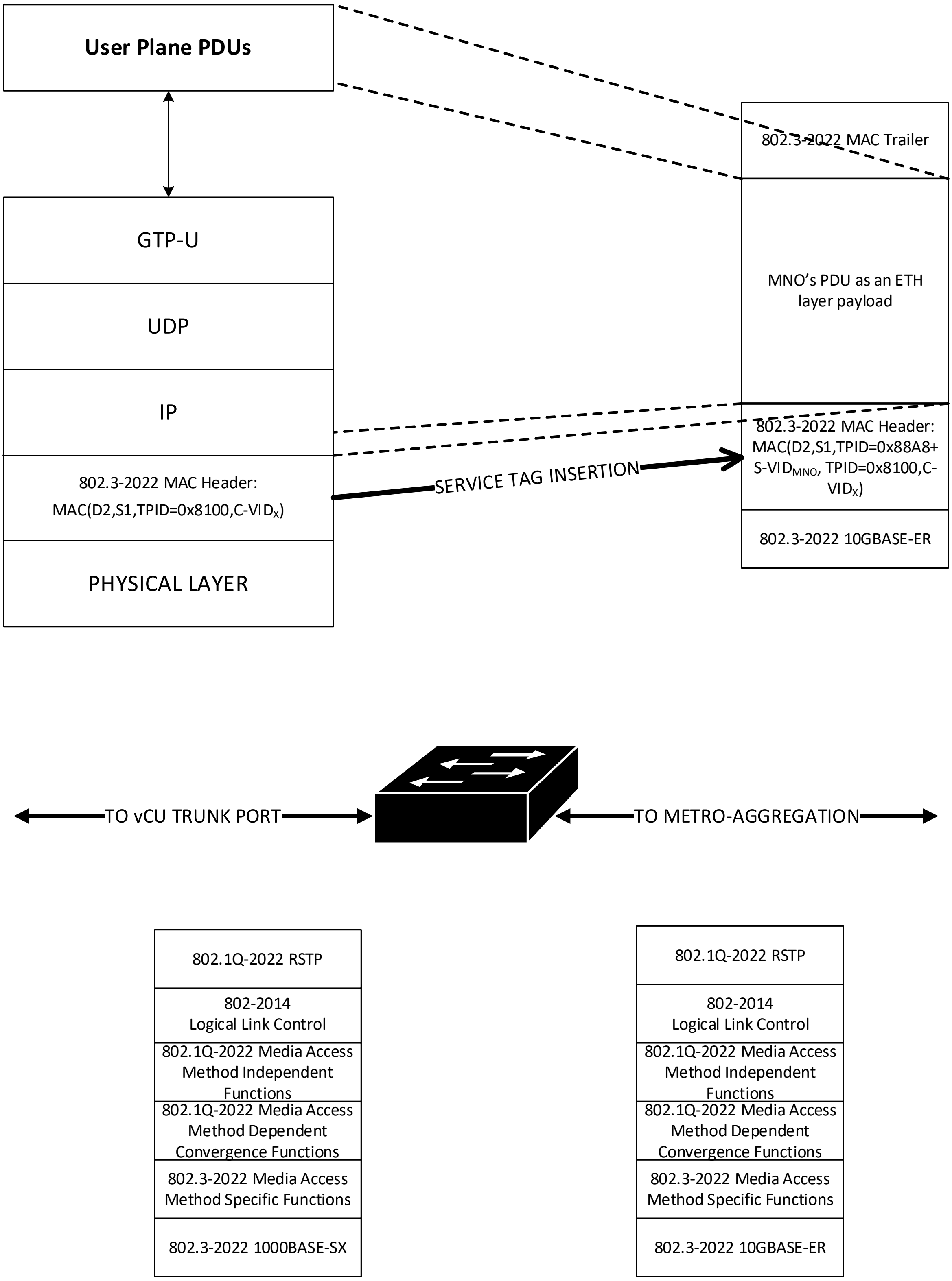


Fig. 102. The recursive nature of transport, in the context of 5G backhaul, using IEEE 802.1 Q Provider Bridging

Here, the LER is the object of attention; again, layers span the range from optical media (layer 0) to MPLS (layer 2+).

To illustrate this second case better, consider again the MNO's operations. As a consequence of the abstraction of all networking entities between the gNB-CU and the UPF, the protocol stack of the PFU at the NG-U [356, Fig. 4.3.1.1–1] must be carried unaltered, collapsing the entire metro-aggregation stack into the equivalent of a layer 2 construct. Here, however, the layer 2 construct is an MPLS artefact (once again: either a link – say a VPWS – or a MAC Bridge – a VPLS) between the gNB-CU and a UPF in the 5GC. The PDUs on the links involved in switching backhauled traffic, are shown in Fig. 104: the MNO is provided with a MEF EVC service with a MEF UNI at the interface between the two physical Ethernet ports linked up within the carrier CSP's[70] access node's real estate.

The specific nature of the MPLS service (itself stacked below an MPLS transport label) is secondary; it could be a VPWS, a VPLS or a BGP EVPN. However, this serves to illustrate the recursive nature of transport: the MNO's Ethernet frame is carried over an MPLS backhaul, which is itself transported over a 10GBASE-ER Ethernet physical layer.

---

[70] A CSP is a carrier when it provides a transport service to another CSP, as is the case here.

Table XXII PB TECHNOLOGY STACK ON METRO ACCESS NODE AGGREGATION SWITCH IN HUB/CO/LE

Layer	Component Type	Descriptor	Comment
Optical media layer	Function	Ring of one pair of single-mode physical optical fibre	G.652
Physical	Function	IEEE 802.3-2022 10GBASE-ER	Wavelength 1550nm, Link <=40km
	Interface	XGMII	PHY-MAC Interface
Link	Function	IEEE 802.3-2022 Media Access Method Specific Functions	Specific IEEE 802 medium access methods, e.g., for 802.3-2022
	Function	IEEE 802.1 Q-2022 Media Access Method Dependent Convergence Functions	Transforms media-access-method independent functions into media-access-method specific functions
	Interface	IEEE 802.1AC-2016 Internal Sublayer Service	Between Media-Access-Method-Dependent-Convergence Functions & Media-Access-Method-Independent Functions
	Function	IEEE 802.1Q-2022 Media Access Method Independent Functions	Two interfaces: (a) EISS towards MAC Relay Entity (b) MS towards LLC
	Interface	IEEE 802.1Q-2022 Enhanced Internal Sublayer Service	Between Media-Access-Method-Independent Functions & MAC Relay Entity Bridge Port
	Function	IEEE 802.1Q-2022 MAC Relay Entity	
	Interface	IEEE 802.1Q-2022 MAC Service	Between Media-Access-Method-Independent Functions & LLC
	Function	IEEE 802-2014 LLC	Maps higher layer protocols to the MSAP according to EtherType or SNAP addresses
	Interface	IEEE 802-2014 LSAP	Between RSTP and LLC
	Function	IEEE 802.1Q-2022 RSTP	Avoids loop formation in the ring

Table XXIII MPLS BACKHAUL TECHNOLOGY STACK ON METRO ACCESS NODE LER (PROVIDER EDGE) IN HUB/CO/LE

Layer	Component Type	Descriptor	Comment
Optical media layer	Function	Ring of one pair of single-mode physical optical fibre	G.652
Physical	Function	IEEE 802.3-2022 10GBASE-ER	Wavelength 1550nm, Link <=40km
	Interface	XGMII	PHY-MAC Interface
Link	Function	IEEE 802.3-2022 Media Access Method Specific Functions	Specific IEEE 802 medium access methods, e.g., for 802.3-2022
	Function	IEEE 802.1 Q-2022 Media Access Method Dependent Convergence Functions	Transforms media-access-method independent functions into media-access-method specific functions
	Interface	IEEE 802.1AC-2016 Internal Sublayer Service	Between Media-Access-Method-Dependent-Convergence Functions & Media-Access-Method-Independent Functions
	Function	IEEE 802.1Q-2022 Media Access Method Independent Functions	
	Interface	IEEE 802.1Q-2022 MAC Service	Between Media-Access-Method-Independent Functions & LLC
	Function	IEEE 802-2014 LLC	Maps higher layer protocols to the MSAP according to EtherType or SNAP addresses
	Interface	IEEE 802-2014 LSAP	Between MPLS and LLC
Above Link, below Network	Function	RFC 3031 MPLS	TR-178-compliant MPLS enabled access node

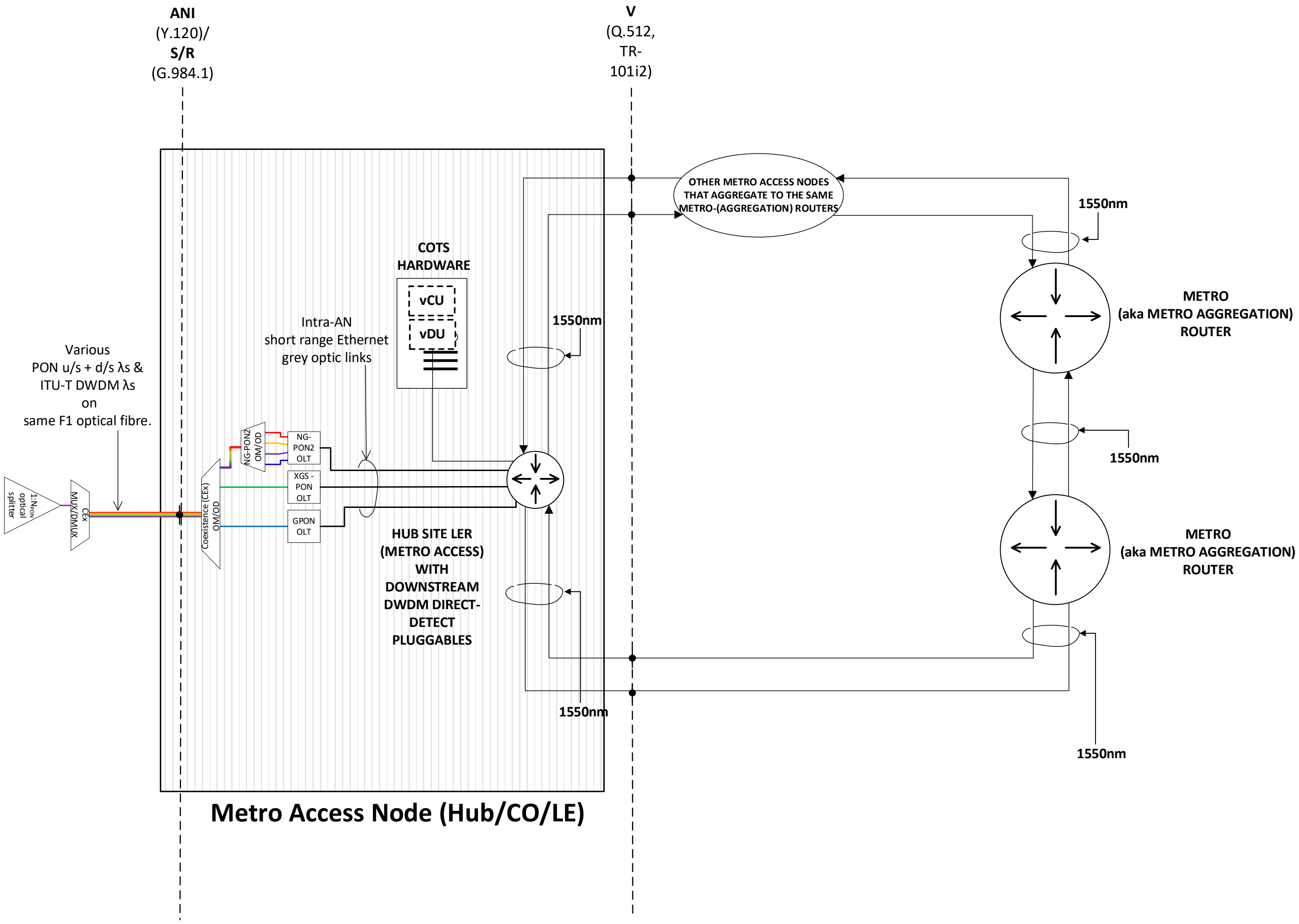


Fig. 103. MPLS backhaul from access node to metro-core over a ring topology

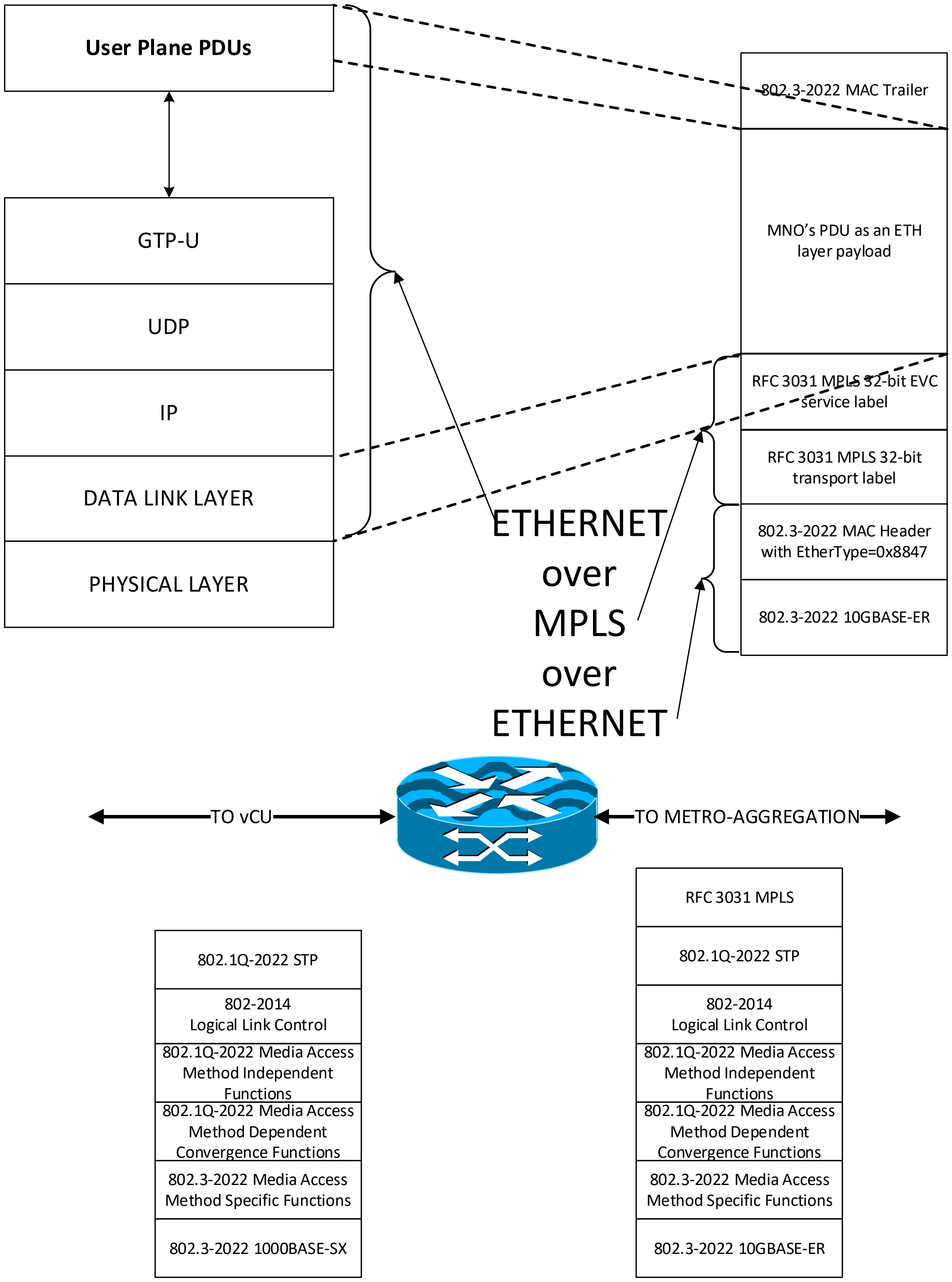


Fig. 104. The recursive nature of transport, in the context of 5G backhaul, using MPLS service, e.g., VPLS, VPWS or BGP EVPN

Chapter 9. Conclusion and further work

The course of this work first addressed the first objective ("identify and solve a particular instance of data inconsistency" – see chapter 3) and subsequently jointly dealt with the second ("reconcile extant architectural paradigms") and the fourth objectives ("develop a baseline for a standardized perspective on current and future metro area networks" – see chapters 3, 4 and 5). Those three chapters (3 – 5) documented a trajectory in modelling that:

- swung first strongly towards the concrete (chapter 3), then
- strongly towards the abstract (chapter 4), and finally
- struck what seems to be a good balance through the superposition of standardized modelling artefacts (RPI-N, IrDI, IaDI and ITU-T G.800 topological components) over schematics of access networks implemented using various wireless and wireline technologies (chapter 5).

Chapters 6 and 7 document the pursuit of the third objective ("identify successful analytical approaches to measurement of energy consumption"). In chapter 6, conclusions were drawn about the prospects of GAL and GALv2 to gain adoption in green path engineering; in chapter 7, a framework for further development of power models of virtualization containers was obtained as a product of application of a novel surveying method (PAD).

Chapter 8 proceeded further down the path of the fourth objective, through the technique of the quantitative and that of the qualitative survey. The surveys have detected the importance of seamless MPLS, the Ethernet PDU at layer 2, simplicity in layers 0 and 1 and suggest that OTN is of secondary importance in the transport plane. There is some interest in ROADMs, yet given the opportunity to use unlit (dark) fibre, this will be preferred. As regards MEC and remote access nodes, while larger CSPs have expressed high interest in customer-proximal MEC nodes, they have not been motivated to move their broadband network gateways out to these deep edge locations.

The techniques used in Chapter 8 emerged out of a three-axis framework, within which the development of method may take place against sound terms of reference:

(a) knowledge about implementation of classical functions,

(b) discernment of major pressures on topological components current at the time during which study is being undertaken, and

(c) discernment of novelty in network functions, along with associated providers of these functions and where they interconnect in the network.

The framework should prove to be a fruitful one. Indeed, two major techniques suggested in chapter 8 – case studies and current or recently-ratified SDO activities – have not been sufficiently exploited. There are several case studies [357], [358], [359] investigating XR optics field deployment, one of which [357] involves TIM (Telecom Italia Mobile). Pertinent SDO activities have been indicated, but not investigated. The ETSI ISG F5G's work is particularly piquing, as it attempts to satisfy all parties playing out the classical technological rift between circuit- and packet-switched networks, as it unfolds in the latest technological arena: {OTN} vs {packet-queues in a DWDM-enabled bandwidth glut}.

Moreover, the G.800 modelling artefacts have not yet been exploited. Trends have been detected and claims have been staked on the basis of the surveys, but they are yet to be mapped out for the span of the metro area network from access node to metro core, in the same way that the transmission media layer has been mapped, and competing alternatives reconciled, for the span from customer up to the access node.

This work is cyclical: the framework is fundamentally cognizant of external pressures for development of the metro area network, and, indeed, telecommunications networks in general. However, future researchers' efforts can benefit from this work as it takes significant steps towards not only establishing a baseline which they can develop, but also equips them with methods which they can employ in their pursuit. I intend to maintain the momentum I have built in bridging the industry – academia divide; I would like to transform the survey into an annual event and update the models in accordance with the results. I have built a few good, mutually beneficial relationships, and would like to pursue the avenue of development of personal relationships at NOG conferences.

At the end of this formative process, I would venture that the ideal destination for all parties, whether CSPs or analysts from academia or industry, would be an SDN platform, informed by implementational models, and complemented by YANG network element models that account for energy consumption. That would go a long way towards meeting the energy analyst's need for a standardized method for data collection.

References

[1] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, ‘Energy Efficiency in the Future Internet: A Survey of Existing Approaches and Trends in Energy-Aware Fixed Network Infrastructures’, *IEEE Commun. Surv. Tutor.*, vol. 13, no. 2, pp. 223–244, 2011, doi: 10.1109/SURV.2011.071410.00073.

[2] F. Idzikowski, L. Chiaraviglio, A. Cianfrani, J. López Vizcaíno, M. Polverini, and Y. Ye, ‘A Survey on Energy-Aware Design and Operation of Core Networks’, *IEEE Commun. Surv. Tutor.*, vol. 18, no. 2, pp. 1453–1499, 2016, doi: 10.1109/COMST.2015.2507789.

[3] R. Maaloul, L. Chaari, and B. Cousin, ‘Energy saving in carrier-grade networks: A survey’, *Comput. Stand. Interfaces*, vol. 55, pp. 8–26, Jan. 2018, doi: 10.1016/j.csi.2017.04.001.

[4] T. Billstein, A. Björklund, and T. Rydberg, ‘Life Cycle Assessment of Network Traffic: A Review of Challenges and Possible Solutions’, *Sustainability*, vol. 13, no. 20, Art. no. 20, Jan. 2021, doi: 10.3390/su132011155.

[5] ‘Energy Reporting’, Lawrence Berkeley National Laboratory. Accessed: Nov. 26, 2022. [Online]. Available: https://ereporting.lbl.gov/

[6] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, ‘Recalibrating global data center energy-use estimates’, *Science*, vol. 367, no. 6481, pp. 984–986, Feb. 2020, doi: 10.1126/science.aba3758.

[7] J. Koomey and E. Masanet, ‘Does not compute: Avoiding pitfalls assessing the Internet’s energy and carbon impacts’, *Joule*, vol. 5, no. 7, pp. 1625–1628, Jul. 2021, doi: 10.1016/j.joule.2021.05.007.

[8] J. Koomey, ‘Does Not Compute: Facts and Fiction about Computing and the Environment’, presented at the iTherm 2021, San Diego, California, Jun. 01, 2021. Accessed: Sep. 08, 2021. [Online]. Available: https://href.li/?https://www.mediafire.com/file/79h3mx3rg1jmkyl/KoomeyslidesforiTherm210601-v5.pdf/file

[9] ‘Blog’, GreeningofStreaming. Accessed: Nov. 28, 2022. [Online]. Available: https://www.greeningofstreaming.org/blog

[10] ‘Home | www.GreeningofStreaming.org’, GreeningofStreaming. Accessed: Nov. 27, 2022. [Online]. Available: https://www.greeningofstreaming.org

[11] D. Schien and C. Preist, ‘A Review of Top-Down Models of Internet Network Energy Intensity’, presented at the ICT for Sustainability 2014 (ICT4S-14), Atlantis Press, Aug. 2014, pp. 87–94. doi: 10.2991/ict4s-14.2014.10.

[12] Dag Lunden, ‘Technical question’, Aug. 07, 2019. Accessed: Nov. 27, 2022. [Online]. Available: https://mail.google.com/mail/u/0/#search/dag+lunden+metro+area/QgrcJHrhvXRdVqnrCPbmFPjTxglWxKrSctL

[13] E.-V. Depasquale and C. Preist, ‘A case for the development of data standards for reporting projections of growth in the Internet’s energy consumption’, presented at the ICT4S, Amsterdam, Aug. 2016.

[14] C. Lange, D. Kosiankowski, R. Weidmann, and A. Gladisch, ‘Energy Consumption of Telecommunication Networks and Related Improvement Options’, *IEEE J. Sel. Top. Quantum Electron.*, vol. 17, no. 2, pp. 285–295, Mar. 2011, doi: 10.1109/JSTQE.2010.2053522.

[15] Dan Kilper and Rodney Tucker, ‘Energy-Efficient Telecommunications’, in *Optical Fiber Telecommunications*, 6th ed., vol. B, 2 vols, I. P. Kaminow, T. Li, and A. E. Wilner, Eds., Elsevier, 2013.

[16] J. Malmodin, D. Lundén, Å. Moberg, G. Andersson, and M. Nilsson, ‘Life Cycle Assessment of ICT’, *J. Ind. Ecol.*, vol. 18, no. 6, pp. 829–845, 2014, doi: 10.1111/jiec.12145.

[17] J. Malmodin and D. Lundén, ‘The energy and carbon footprint of the ICT and E&M sector in Sweden 1990-2015 and beyond’, presented at the ICT for Sustainability 2016, Atlantis Press, Aug. 2016, pp. 209–218. doi: 10.2991/ict4s-16.2016.25.

[18] ‘Broadband Portal - OECD’. Accessed: Dec. 03, 2022. [Online]. Available: https://www.oecd.org/sti/broadband/broadband-statistics/

[19] W. B. Norton, *The 2014 Internet Peering Playbook: Connecting to the Core of the Internet*. DrPeering Press, 2014.

[20] L. Andersson and T. Madsen, ‘Provider Provisioned Virtual Private Network (VPN) Terminology’, Internet Engineering Task Force, Request for Comments RFC 4026, Mar. 2005. doi: 10.17487/RFC4026.

[21] R. D. Doverspike, K. K. Ramakrishnan, and C. Chase, ‘Structural Overview of ISP Networks’, in *Guide to Reliable Internet Services and Applications*, C. R. Kalmanek, S. Misra, and Y. (Richard) Yang, Eds., in Computer Communications and Networks. , London: Springer, 2010, pp. 19–93. doi: 10.1007/978-1-84882-828-5_2.

[22] ‘Modular Headend Architecture v2 Technical Report’, Cable Television Laboratories, Inc, CM-TR-MHAv2-V01-150615, Jun. 2015.

[23] E.-V. Depasquale, S. Zammit, F. Davoli, and R. Bolla, ‘An Implementational Model of Telecommunications Networks for the Energy Analyst’, in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Jul. 2018, pp. 336–344. doi: 10.1109/Cybermatics_2018.2018.00085.

[24] ‘Open XR Forum’, Open XR Forum. Accessed: Dec. 04, 2022. [Online]. Available: https://www.openxrforum.org

[25] J. Bäck *et al.*, ‘CAPEX Savings Enabled by Point-to-Multipoint Coherent Pluggable Optics Using Digital Subcarrier Multiplexing in Metro Aggregation Networks’, in *2020 European Conference on Optical Communications (ECOC)*, Dec. 2020, pp. 1–4. doi: 10.1109/ECOC48923.2020.9333233.

[26] E.-V. Depasquale, ‘Corrigenda: ITU-T G.8310 (Figs. 8-1 - 8-2)’, Oct. 27, 2021. [Online]. Available: https://drive.google.com/file/d/1rPG9fWmmubg4iclcqucJwhsGFRPmoLSQ/view?usp=sharing

[27] E.-V. Depasquale, ‘Corrigendum - G.805 (Figs 7, 8)’, Nov. 01, 2021. [Online]. Available: https://drive.google.com/file/d/1rOYHHal2aceDhjrbeC8MfEVwt_kT2iky/view?usp=sharing

[28] E.-V. Depasquale, ‘Corrigendum - Y.110’, Sep. 14, 2021. [Online]. Available: https://drive.google.com/file/d/1rGwy0Ve86JVCmHmRPTzcbceWZwe08hzS/view?usp=sharing

[29] E.-V. Depasquale, ‘Corrigendum: G.800 (para 3.2.8)’, Oct. 26, 2021. [Online]. Available: https://drive.google.com/file/d/1r76XlsH98e3QPVGcYqwxgEvQw_LkVqQk/view?usp=sharing

[30] S. Shew, ‘Updates to ITU-T G.800, G.805, and G.8310’, Dec. 17, 2021. [Online]. Available: https://drive.google.com/open?id=1cEudgw0iE_Gh-bOb4hbG8EP9WqriMx8L&usp=drive_fs

[31] ‘Architecture of the metro transport network - Corrigendum 1’. Rec. ITU-T G.8310 Corrigendum 1, International Telecommunication Union. [Online]. Available: https://www.itu.int/rec/T-REC-G.8310-202202-I!Cor1/en

[32] ‘Generic functional architecture of transport networks - Corrigendum 1’. Rec. ITU-T G.805 Corrigendum 1, International Telecommunication Union. [Online]. Available: https://www.itu.int/rec/T-REC-G.805-202202-I!Cor1/en

[33] ‘Implementers’ guide for Recommendation ITU-T Y.110 “Global Information Infrastructure principles and framework architecture”’. Rec. ITU-T Y.110 Implementers’ guide, International Telecommunication Union, Dec. 2021.

[34] ‘Unified functional architecture of transport networks - Corrigendum 1’. Rec. ITU-T G.800 Corrigendum 1, International Telecommunication Union, Feb. 2022. [Online]. Available: https://www.itu.int/rec/T-REC-G.800-202202-I!Cor1/en

[35] R. Shackelford *et al.*, ‘Computing Curricula 2005: The Overview Report’, in *Proceedings of the 37th SIGCSE technical symposium on Computer science education*, in SIGCSE ’06. Houston, Texas, USA: Association for Computing Machinery, Mar. 2006, pp. 456–457. doi: 10.1145/1121341.1121482.

[36] L. Hilty and W. Lohmann, ‘An Annotated Bibliography of Conceptual Frameworks in ICT for Sustainability’, in *Hilty, Lorenz; Lohmann, Wolfgang (2013). An Annotated Bibliography of Conceptual Frameworks in ICT for Sustainability. In: ICT4S – First International Conference on Information and Communication Technologies for Sustainability, Zürich, 12 February 2013 - 14 February 2013. ETH E-Collection, 288-300.*, L. Hilty, B. Aebischer, G. Andersson, and W. Lohmann, Eds., Zürich: ETH E-Collection, Feb. 2013, pp. 288–300. doi: 10.5167/uzh-84887.

[37] S. Murugesan and G. R. Gangadharan, *Harnessing Green IT: Principles and Practices*. J. Wiley & Sons, 2012.

[38] C. DiSalvo, P. Sengers, and H. Brynjarsdóttir, ‘Mapping the landscape of sustainable HCI’, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, in CHI ’10. New York, NY, USA: Association for Computing Machinery, Apr. 2010, pp. 1975–1984. doi: 10.1145/1753326.1753625.

[39] F. Berkhout and J. Hertin, ‘Impacts of information and communication technologies on environmental sustainability: Speculations and evidence’, OECD, Brighton, May 2001.

[40] K. Ishii, J. Kurumida, K. Sato, T. Kudoh, and S. Namiki, ‘Unifying Top-Down and Bottom-Up Approaches to Evaluate Network Energy Consumption’, *J. Light. Technol.*, vol. 33, no. 21, pp. 4395–4405, Nov. 2015, doi: 10.1109/JLT.2015.2469145.

[41] ‘2016 Annual Report Voluntary Agreement for Ongoing Improvement to the Energy Efficiency of Set-Top Boxes’, 2017, Accessed: Jul. 04, 2020. [Online]. Available: https://www.energy-efficiency.us/library/pdf/STB2016AnnualReport.pdf

[42] ‘Cisco visual networking index: Forecast and methodology, 2014–2019’, Cisco Systems, San Jose, CA, 2015.

[43] ‘Cisco visual networking index: Forecast and methodoolgy, 2015-2020’, Cisco Systems, San Jose, CA, 2016.

[44] ‘Cisco visual networking index: Forecast and methodology, 2016–2021’, Cisco Systems, San Jose, CA, 2017.

[45] ‘Cisco visual networking index: Forecast and Trends, 2017-2022’, Cisco Systems, San Jose, CA, 2019.

[46] ‘Metro network traffic growth: an architecture impact study’, Alcatel-Lucent, 2013.

[47] R. Bolla, F. Davoli, R. Bruschi, K. Christensen, F. Cucchietti, and S. Singh, ‘The potential impact of green technologies in next-generation wireline networks: Is there room for energy saving optimization?’, *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 80–86, Aug. 2011, doi: 10.1109/MCOM.2011.5978419.

[48] ‘Reducing the Net Energy Consumption in Energy Communications Networks by up to 98% by 2020’, GreenTouch Consortium, Jun. 2015. Accessed: May 10, 2023. [Online]. Available: https://www.bell-labs.com/greentouch/uploads/documents/GreenTouch_Green_Meter_Final_Results_18_June_2015.pdf

[49] C. Lange, ‘Energy-related Aspects in Backbone Networks.’, in *Proceedings of 35th European Conference on Optical Communication (ECOC 2009)*, Vienna, Austria, Sep. 2009.

[50] ‘G.W.A.T.T.: Global “What if” Analyzer of neTwork energy consumpTion’, GreenTouch Consortium, 2015. Accessed: Jan. 23, 2016. [Online]. Available: http://gwatt.net/intro/2

[51] ‘G.W.A.T.T.: Global “What if” Analyzer of neTwork energy consumpTion’, GreenTouch Consortium, 2015. Accessed: Jan. 23, 2016. [Online]. Available: http://gwatt.net/intro/4

[52] ‘G.W.A.T.T.: Global “What if” Analyzer of neTwork energy consumpTion’, GreenTouch Consortium, 2015. Accessed: Jan. 23, 2016. [Online]. Available: http://gwatt.net/intro/5

[53] ‘G.W.A.T.T.: Global “What if” Analyzer of neTwork energy consumpTion’, GreenTouch Consortium, 2015. Accessed: Jan. 23, 2016. [Online]. Available: http://gwatt.net/intro/6

[54] ‘Global Internet Phenomena Report’, Sandvine, 2013. Accessed: Jul. 04, 2020. [Online]. Available: https://www.sandvine.com/hubfs/downloads/archive/2013-2h-global-internet-phenomena-report.pdf

[55] ‘Global Internet Phenomena - Africa, Middle East & North America’, Dec. 2015. Accessed: Jul. 04, 2020. [Online]. Available: https://www.sandvine.com/hubfs/downloads/archive/2015-global-internet-phenomena-report-africa-middle-east-and-north-america.pdf

[56] ‘Global Internet Phenomena Report’, Sandvine, Sep. 2019.

[57] J. Brodkin, ‘Verizon seeks payment for carrying Netflix traffic, WSJ reports’, Ars Technica. Accessed: May 10, 2023. [Online]. Available: https://arstechnica.com/information-technology/2014/02/verizon-seeks-payment-for-carrying-netflix-traffic-wsj-reports/

[58] ‘Comcast vs. Netflix: Is this really about Net neutrality?’, CNET. Accessed: May 10, 2023. [Online]. Available: https://www.cnet.com/tech/services-and-software/comcast-vs-netflix-is-this-really-about-net-neutrality/

[59] W. Team, ‘Why Netflix Chose NGINX as the Heart of Its CDN’, NGINX. Accessed: May 10, 2023. [Online]. Available: https://www.nginx.com/blog/why-netflix-chose-nginx-as-the-heart-of-its-cdn/

[60] U. Lee, I. Rimac, and V. Hilt, ‘Greening the internet with content-centric networking’, in *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*, in e-Energy ’10. Passau, Germany: Association for Computing Machinery, Apr. 2010, pp. 179–182. doi: 10.1145/1791314.1791342.

[61] K. Guan, G. Atkinson, D. C. Kilper, and E. Gulsen, ‘On the Energy Efficiency of Content Delivery Architectures’, in *2011 IEEE International Conference on Communications Workshops (ICC)*, Jun. 2011, pp. 1–6. doi: 10.1109/iccw.2011.5963557.

[62] N. Choi, K. Guan, D. C. Kilper, and G. Atkinson, ‘In-network caching effect on optimal energy consumption in content-centric networking’, in *2012 IEEE International Conference on Communications (ICC)*, Jun. 2012, pp. 2889–2894. doi: 10.1109/ICC.2012.6364320.

[63] N. I. Osman, T. El-Gorashi, and J. M. H. Elmirghani, ‘Reduction of energy consumption of Video-on-Demand services using cache size optimization’, in *2011 Eighth International Conference on Wireless and Optical Communications Networks*, May 2011, pp. 1–5. doi: 10.1109/WOCN.2011.5872923.

[64] J. Llorca *et al.*, ‘Dynamic in-network caching for energy efficient content delivery’, in *2013 Proceedings IEEE INFOCOM*, Apr. 2013, pp. 245–249. doi: 10.1109/INFCOM.2013.6566772.

[65] U. Mandal, P. Chowdhury, C. Lange, A. Gladisch, and B. Mukherjee, ‘Energy-efficient networking for content distribution over telecom network infrastructure’, *Opt. Switch. Netw.*, vol. 10, no. 4, pp. 393–405, Nov. 2013, doi: 10.1016/j.osn.2013.06.003.

[66] R. Modrzejewski *et al.*, ‘Energy efficient content distribution in an ISP network’, in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2013, pp. 2859–2865. doi: 10.1109/GLOCOM.2013.6831508.

[67] N. Abji, A. Tizghadam, and A. Leon-Garcia, ‘Energy efficient content delivery in service provider networks with content caching’, in *2015 IEEE Online Conference on Green Communications (OnlineGreenComm)*, Nov. 2015, pp. 23–29. doi: 10.1109/OnlineGreenCom.2015.7387374.

[68] M. Savi, O. Ayoub, F. Musumeci, Z. Li, G. Verticale, and M. Tornatore, ‘Energy-efficient caching for Video-on-Demand in Fixed-Mobile Convergent networks’, in *2015 IEEE Online Conference on Green Communications (OnlineGreenComm)*, Nov. 2015, pp. 17–22. doi: 10.1109/OnlineGreenCom.2015.7387373.

[69] C. Jayasundara, A. Nirmalathas, E. Wong, and C. Chan, ‘Improving Energy Efficiency of Video on Demand Services’, *IEEEOSA J. Opt. Commun. Netw.*, vol. 3, no. 11, pp. 870–880, Nov. 2011, doi: 10.1364/JOCN.3.000870.

[70] R. Fratini, M. Savi, G. Verticale, and M. Tornatore, ‘Using replicated video servers for VoD traffic offloading in integrated metro/access networks’, in *2014 IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 3438–3443. doi: 10.1109/ICC.2014.6883853.

[71] E. D. Pascale, D. B. Payne, and M. Ruffini, ‘Bandwidth and energy savings of locality-aware P2P Content Distribution in next-generation PONs’, in *2012 16th International Conference on Optical Network Design and Modelling (ONDM)*, Apr. 2012, pp. 1–6. doi: 10.1109/ONDM.2012.6210215.

[72] ‘Future networks: Objectives and design goals’. Rec. ITU-T Y.3001, International Telecommunication Union, Sophia Antipolis, France, May 2011.

[73] Y.-D. Lin, D. Pitt, D. Hausheer, E. Johnson, and Y.-B. Lin, ‘Software-Defined Networking: Standardization for Cloud Computing’s Second Wave’, *Computer*, vol. 47, no. 11, pp. 19–21, Nov. 2014, doi: 10.1109/MC.2014.329.

[74] R. Bolla, R. Bruschi, F. Davoli, and C. Lombardo, ‘Fine-Grained Energy-Efficient Consolidation in SDN Networks and Devices’, *IEEE Trans. Netw. Serv. Manag.*, vol. 12, no. 2, pp. 132–145, Jun. 2015, doi: 10.1109/TNSM.2015.2431074.

[75] W. Stallings, *Foundations of modern networking: SDN, NFV, QoE, IoT and Cloud.* Addison-Wesley Professional, 2015.

[76] S. Shenker, M. Casado, T. Koponen, and N. McKeown, ‘The future of networking, and the past of protocols’, presented at the Open Networking Summit, 2011.

[77] T. D. Nadeau and K. Gray, *SDN: Software Defined Networks: An authoritative review of network programmability technologies.* O'Reilly Media, Inc., 2013.

[78] V. Ganti, V. Lubsey, M. Shekhar, and C. Swan, 'Software Defined Networking Master Usage Model'. Open Data Center Alliance, 2013. Accessed: May 11, 2016. [Online]. Available: https://opendatacenteralliance.org/wp-content/uploads/Software_Defined_Networking_Master_Usage_Model_Rev1.0.pdf

[79] 'Pervasive Mobile Virtual Services', Expert Advisory Group of the European Technology Platform Networld 2020, Jul. 2016. [Online]. Available: https://www.networld2020.eu/wp-content/uploads/2014/02/SRIA_final.pdf

[80] '5G Wireless Wireline Convergence Architecture', TR-470 Issue 2, Broadband Forum, Mar. 2022. Accessed: Oct. 03, 2020. [Online]. Available: https://www.broadband-forum.org/technical/download/TR-470.pdf

[81] 'Network Functions Virtualisation (NFV); Architectural Framework', Group Specification ETSI GS NFV 002 V1.2.1, Dec. 2014. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV/001_099/002/01.02.01_60/gs_NFV002v010201p.pdf

[82] R. Bolla, R. Bruschi, F. Davoli, C. Lombardo, J. F. Pajo, and O. R. Sanchez, 'The dark side of network functions virtualization: A perspective on the technological sustainability', in *IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7. doi: 10.1109/ICC.2017.7997129.

[83] 'Network Functions Virtualisation (NFV); Use Cases', ETSI, RGR/NFV-001ed121, May 2017.

[84] W. Shi, 'Making sense of the telco cloud', Telecoms.com, May 2020. Accessed: Jun. 27, 2020. [Online]. Available: https://telecoms.com/opinion/making-sense-of-the-telco-cloud/

[85] 'CNF friendly networking for Telco/Edge Kubernetes platforms', Sep. 2020, Accessed: Sep. 28, 2020. [Online]. Available: https://www.brighttalk.com/webcast/12229/438035/cnf-friendly-networking-for-telco-edge-kubernetes-platforms

[86] J. Crawshaw, 'Deconstructing the Telco Cloud', *Light Read.*, Apr. 2019, Accessed: Jul. 25, 2020. [Online]. Available: https://www.lightreading.com/cloud/deconstructing-the-telco-cloud/a/d-id/753874

[87] S. E. E. Ayoubi *et al.*, 'Refined scenarios and requirements, consolidated use cases, and qualitative techno-economic feasibility assessment', D1.1, Jan. 2016. Accessed: Jun. 29, 2020. [Online]. Available: https://metis-ii.5g-ppp.eu/wp-content/uploads/deliverables/METIS-II_D1.1_v1.0.pdf

[88] J. Muthurajan, 'Why Use Cloud-Native Containerized Network Functions (CNFs)?', Intel, White paper. Accessed: Jul. 25, 2020. [Online]. Available: https://www.intel.com/content/www/uk/en/communications/why-containers-and-cloud-native-functions-paper.html

[89] R. Vettor and S. Ardalis Smith, 'Architecting Cloud Native .NET Applications for Azure', Oct. 2020, [Online]. Available: https://docs.microsoft.com/en-us/dotnet/architecture/cloud-native/

[90] M. Logan, 'CNCF F Technical Oversight Committee (TOOC)', *GitHub*, Mar. 2021, [Online]. Available: https://github.com/cncf/toc

[91] 'Environmental management - Life cycle assessment - Principles and framework'. ISO 14040:2006, International Organization for Standardization, Jul. 2006. Accessed: Jul. 27, 2020. [Online]. Available: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/74/37456.html

[92] 'Environmental management - Life cycle assessment - Requirements and guidelines'. ISO 14044:2006, International Organization for Standardization, Jul. 2006. Accessed: Nov. 27, 2022. [Online]. Available: https://www.iso.org/standard/38498.html

[93] M. A. Curran, *Goal and Scope Definition in Life Cycle Assessment*. Springer Netherlands, 2016.

[94] D. Schien, P. Shabajee, M. Yearworth, and C. Preist, 'Modeling and Assessing Variability in Energy Consumption During the Use Stage of Online Multimedia Services', *J. Ind. Ecol.*, vol. 17, no. 6, pp. 800–813, 2013, doi: 10.1111/jiec.12065.

[95] 'Environmental management - Life cycle assessment - Requirements and guidelines for organizational life cycle assessment'. ISO 14072:2014, International Organization for Standardization, Dec. 2014. [Online]. Available: https://www.iso.org/standard/61104.html

[96] L. Shi, K. J. Mach, S. Suh, and A. Brandt, ‘Functionality-based life cycle assessment framework: An information and communication technologies (ICT) product case study’, *J. Ind. Ecol.*, vol. 26, no. 3, pp. 782–800, 2022, doi: 10.1111/jiec.13240.

[97] R. Heijungs and S. Suh, *The Computational Structure of Life Cycle Assessment*. Springer Science & Business Media, 2002.

[98] S. Suh and G. Huppes, ‘Methods for Life Cycle Inventory of a product’, *J. Clean. Prod.*, vol. 13, no. 7, pp. 687–697, Jun. 2005, doi: 10.1016/j.jclepro.2003.04.001.

[99] A. Stavdas, *Core and Metro Networks*. John Wiley & Sons, 2010.

[100] ‘Global Information Infrastructure principles and framework architecture’. Rec. ITU-T Y.110, International Telecommunication Union, Jun. 12, 1998. [Online]. Available: https://www.itu.int/rec/T-REC-Y.110/en

[101] ‘Global Information Infrastructure (GII): Reference points for interconnection framework’. Rec. ITU-T Y.140, International Telecommunication Union, Nov. 2000. [Online]. Available: https://www.itu.int/rec/T-REC-Y.140-200011-I/en

[102] C. Filsfils, S. Previdi, L. Ginsberg, B. Decraene, S. Litkowski, and R. Shakir, ‘Segment Routing Architecture’, Internet Engineering Task Force, Request for Comments RFC 8402, Jul. 2018. doi: 10.17487/RFC8402.

[103] J. M. Halpern and C. Pignataro, ‘Service Function Chaining (SFC) Architecture’, Internet Engineering Task Force, Request for Comments RFC 7665, Oct. 2015. doi: 10.17487/RFC7665.

[104] P. Matthews, I. van Beijnum, and M. Bagnulo, ‘Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers’, Internet Engineering Task Force, Request for Comments RFC 6146, Apr. 2011. doi: 10.17487/RFC6146.

[105] ‘G.W.A.T.T. (Global “What if” Analyzer of NeTwork Energy ConsumpTion)’, Bell Labs application able to measure the impact of technologies like SDN & NFV on network energy consumption., 2015. [Online]. Available: https://media-bell-labs-com.s3.amazonaws.com/pages/20150114_1907/GWATT_WhitePaper.pdf

[106] A. P. Bianzino, C. Chaudet, D. Rossi, and J. Rougier, ‘A Survey of Green Networking Research’, *IEEE Commun. Surv. Tutor.*, vol. 14, no. 1, pp. 3–20, 2012, doi: 10.1109/SURV.2011.113010.00106.

[107] V. C. Coroama and L. M. Hilty, ‘Assessing Internet energy intensity: A review of methods and results’, *Environ. Impact Assess. Rev.*, vol. 45, pp. 63–68, Feb. 2014, doi: 10.1016/j.eiar.2013.12.004.

[108] K. Grobe, J.-P. Elbers, and S. Neidlinger, ‘Broadband Access Networks’, in *Broadband Access Networks: Technologies and Deployments*, A. Shami, M. Maier, and C. Assi, Eds., in Optical Networks. , Boston, MA: Springer US, 2009, pp. 353–372. doi: 10.1007/978-0-387-92131-0_17.

[109] Y. Rekhter and E. C. Rosen, ‘BGP/MPLS IP Virtual Private Networks (VPNs)’, Internet Engineering Task Force, Request for Comments RFC 4364, Feb. 2006. doi: 10.17487/RFC4364.

[110] Y. Rekhter and K. Kompella, ‘Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling’, Internet Engineering Task Force, Request for Comments RFC 4761, Jan. 2007. doi: 10.17487/RFC4761.

[111] ‘Metro Ethernet Network Architecture Framework - Part 1: Generic Framework’. MEF 4, Metro Ethernet Forum, May 2004. [Online]. Available: https://www.mef.net/resources/mef-4-metro-ethernet-network-architecture-framework-part-1-generic-framework/

[112] A. Herzog, ‘The Coming Carrier Network Infrastructure: A Way Different Landscape’, Alcatel-Lucent, 2007.

[113] M. Forzati *et al.*, ‘Integrated OASE Results Overview’, CORDIS, Project Deliverable Open Access Seamless Evolution Deliverable 8.5, Oct. 2013. [Online]. Available: https://cordis.europa.eu/docs/projects/cnect/5/249025/080/deliverables/001-OASED85WP8DTAG15102013V20.pdf

[114] ‘Gigabit Ethernet CWDM 80km SFP Optical Transceiver | Coherent Corp.’ Accessed: May 14, 2023. [Online]. Available: https://ii-vi.com/product/gigabit-ethernet-cwdm-80km-sfp-optical-transceiver/

[115] ‘Datacom Telecom Network Diagram - Finisar’. Accessed: Apr. 03, 2016. [Online]. Available: https://www.finisar.com/sites/default/files/resources/datacom_telecom_diagrams_handout_-_2015.pdf.

[116] I. Tomkos, 'Metropolitan area optical networks', *IEEE Circuits and Devices Magazine*, vol. 19, no. 4, pp. 24–30, Jul. 2003.

[117] 'Cisco ME 4924-10GE Ethernet Aggregation Switch'. Accessed: Jul. 15, 2023. [Online]. Available: https://www.cisco.com/c/en/us/products/collateral/switches/me-4900-series-ethernet-switches/product_data_sheet0900aecd8052f36b.html

[118] R. Chundury, 'Mobile broadband backhaul: Addressing the challenge', *Ericsson Review*, vol. 2008, no. 3, pp. 4–9, 2008.

[119] A. Gumaste and T. Antony, *DWDM Network Designs and Engineering Solutions*. Cisco Press, 2003.

[120] P. Castoldi *et al.*, 'Design of Reliable Metro Core Networks', in *2007 9th International Conference on Transparent Optical Networks*, Jul. 2007, pp. 130–133. doi: 10.1109/ICTON.2007.4296263.

[121] M. Maier, *Optical Switching Networks*. Cambridge University Press, 2008.

[122] 'Amsterdam Internet Exchange News'. Accessed: Apr. 05, 2016. [Online]. Available: https://ams-ix.net/newsitems/23

[123] 'Upscaling the metro - Deploying the 7950 XRS in the metro core'. Accessed: Feb. 01, 2016. [Online]. Available: http://resources.alcatel-lucent.com/asset/171550

[124] A. Jansen, 'The Case for the Metro Core', TechZine. Accessed: Apr. 16, 2016. [Online]. Available: https://techzine.alcatel-lucent.com/case-metro-core

[125] P. Kwon, '2015 Japan - Telecoms, Mobile and Broadband - BuddeComm'. Accessed: May 17, 2023. [Online]. Available: https://www.budde.com.au/Research/2015-Japan-Telecoms-Mobile-and-Broadband

[126] A. S. G. Andrae, 'Total consumer power consumption forecast', presented at the Nordic Digital Business Summit, Helsinki, 2017.

[127] 'General principles and general reference model for Next Generation Networks.' Rec. ITU-T Y.2011, International Telecommunication Union, Oct. 07, 2004. [Online]. Available: https://www.itu.int/rec/T-REC-Y.2011/en

[128] 'Global Information Infrastructure terminology: Terms and definitions'. Rec. ITU-T Y.101, International Telecommunication Union, Mar. 2000. [Online]. Available: https://www.itu.int/rec/T-REC-Y.101/en

[129] 'General overview of NGN'. Rec. ITU-T Y.2001, International Telecommunication Union, Dec. 17, 2004. [Online]. Available: https://www.itu.int/rec/T-REC-Y.2001/en

[130] 'Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture'. ES 282 001, European Telecommunications Standards Institute, Sophia Antipolis, France, Sep. 09, 2009. [Online]. Available: https://portal.etsi.org/webapp/workprogram/Report_WorkItem.asp?WKI_ID=28410

[131] 'Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Technical Report on NGN National IP Interconnection'. TR 180 003, European Telecommunication, Sophia Antipolis, France, Apr. 11, 2012. [Online]. Available: https://portal.etsi.org/webapp/workprogram/Report_WorkItem.asp?WKI_ID=30995

[132] European Telecommunications Standards Institute, 'Next Generation Networks - Our Role & Activities'. Accessed: Mar. 03, 2018. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/past-work/next-generation-networks

[133] 'Broadband Multi-Service Architecture and Framework Requirements'. TR-144, Broadband Forum, Aug. 2007. [Online]. Available: https://www.broadband-forum.org/technical/download/TR-144.pdf

[134] 'Multi-service Broadband Network Functional Modules and Architectures'. TR-145, Broadband Forum, Nov. 2012. [Online]. Available: https://www.broadband-forum.org/technical/download/TR-145.pdf

[135] D. Weller and B. Woodcock, 'Internet Traffic Exchange: Market Developments and Policy Challenges', OECD, Paris, Jan. 2013. doi: 10.1787/5k918gpt130q-en.

[136] 'ISDN user-network interfaces - reference configurations'. Rec. ITU-T I.411, International Telecommunication Union, Mar. 1993. [Online]. Available: https://www.itu.int/rec/T-REC-I.411/en

[137] 'Global information infrastructure scenario methodology Annex A: Examples of use'. Rec. ITU-T Y.120 Annex A, International Telecommunication Union, Feb. 1999. [Online]. Available: https://www.itu.int/rec/T-REC-Y.120-199902-I!AnnA/en

[138] D. Meyer and R. Bush, 'Some Internet Architectural Guidelines and Philosophy', Internet Engineering Task Force, Request for Comments RFC 3439, Dec. 2002. doi: 10.17487/RFC3439.

[139] 'Definitions of terms related to quality of service'. Rec. ITU-T E.800, International Telecommunication Union, Sep. 23, 2008. [Online]. Available: https://www.itu.int/rec/T-REC-E.800-200809-I/en

[140] D. Schien, P. Shabajee, S. G. Wood, and C. Preist, 'A model for green design of online news media services', in *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, Rio de Janeiro, Brazil: ACM Press, 2013, pp. 1111–1122. doi: 10.1145/2488388.2488485.

[141] W. Vereecken *et al.*, 'Power consumption in telecommunication networks: overview and reduction strategies', *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 62–69, Jun. 2011, doi: 10.1109/MCOM.2011.5783986.

[142] D. Schien, V. C. Coroama, L. M. Hilty, and C. Preist, 'The Energy Intensity of the Internet: Edge and Core Networks', in *ICT Innovations for Sustainability*, L. M. Hilty and B. Aebischer, Eds., Cham: Springer International Publishing, 2015, pp. 157–170. doi: 10.1007/978-3-319-09228-7_9.

[143] D. C. Kilper *et al.*, 'Power Trends in Communication Networks', *IEEE J. Sel. Top. Quantum Electron.*, vol. 17, no. 2, pp. 275–284, Mar. 2011, doi: 10.1109/JSTQE.2010.2074187.

[144] L. Paraschis, O. Gerstel, and M. Y. Frankel, 'Metro networks: Services and technologies', in *Optical Fiber Telecommunications V B: Systems and Networks*, 5th ed., vol. B, Cambridge, MA: Elsevier Inc., 2008, pp. 477–509.

[145] 'Broadband Remote Access Server (BRAS) Requirements Document'. TR-092, Broadband Forum, Aug. 2004.

[146] 'A Border Gateway Protocol 4 (BGP-4)', Internet Engineering Task Force, Request for Comments RFC 1771, Mar. 1995. doi: 10.17487/RFC1771.

[147] J. Davis and K. Morgan, 'Interconnection 101', 451 Research, 2015.

[148] J. Malmodin, D. Lundén, M. Nilsson, and G. Andersson, 'LCA of data transmission and IP core networks', in *2012 Electronics Goes Green 2012+*, Sep. 2012, pp. 1–6.

[149] 'Migration to Ethernet-Based Broadband Aggregation', TR-101 Issue 2, Broadband Forum, Jul. 2011.

[150] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, and A. Neal, 'Mobile-edge computing introductory technical white paper.', mobile-edge computing (MEC) industry initiative, White paper.

[151] 'Information Technology - Generic digital audio-visual systems - Part 1: System reference models and scenarios'. ISO/IEC 16500-1, International Organization for Standardization, Dec. 1999.

[152] 'DOCSIS 3.1 Physical Layer Specification'. CM-SP-PHYv3.1, CableLabs, Apr. 19, 2023. [Online]. Available: https://www.cablelabs.com/specifications/CM-SP-PHYv3.1

[153] 'Protocols at the U Interface for Accessing Data Networks using ATM/DSL'. TR-043, Broadband Forum, Aug. 2001. [Online]. Available: https://www.broadband-forum.org/technical/download/TR-043.pdf

[154] 'Framework Recommendation on functional access networks (AN) - Architecture and functions, access types, management and service node aspects'. Rec. ITU-T G.902, International Telecommunication Union, Nov. 1995. Accessed: Jan. 11, 2022. [Online]. Available: https://www.itu.int/rec/T-REC-G.902/en

[155] 'Using GPON Access in the context of TR-101'. TR-156 Issue 4, Broadband Forum, Nov. 2017. [Online]. Available: https://www.broadband-forum.org/pdfs/tr-156-4-0-0.pdf

[156] J. Lorincz, A. Capone, and J. Wu, 'Greener, Energy-Efficient and Sustainable Networks: State-Of-The-Art and New Trends', *Sensors*, vol. 19, no. 22, Art. no. 22, Jan. 2019, doi: 10.3390/s19224864.

[157] M. Usama and M. Erol-Kantarci, 'A Survey on Recent Trends and Open Issues in Energy Efficiency of 5G', *Sensors*, vol. 19, no. 14, Art. no. 14, Jan. 2019, doi: 10.3390/s19143126.

[158] P. Cong, J. Zhou, L. Li, K. Cao, T. Wei, and K. Li, 'A Survey of Hierarchical Energy Optimization for Mobile Edge Computing', *ACM Comput. Surv. CSUR*, Apr. 2020, doi: 10.1145/3378935.

[159] ITU-T Y.110, 'Global Information Infrastructure Principles and Framework Architecture', International Telecommunication Union, Jun. 1998.

[160] 'Vocabulary of terms for ISDNs'. Rec. ITU-T I.112, International Telecommunication Union, Mar. 1993. [Online]. Available: https://www.itu.int/rec/T-REC-I.112/en

[161] 'Generic functional architecture of transport networks'. Rec. ITU-T G.805, International Telecommunication Union, Mar. 2000. [Online]. Available: https://www.itu.int/rec/T-REC-G.805/en

[162] J. P. Clark, N. Barcet, and F. Zdarsky, 'Red Hat OpenShift's road to the network edge', Jul. 22, 2020. Accessed: Aug. 06, 2020. [Online]. Available: https://on24static.akamaized.net/event/24/94/44/2/rt/1/documents/resourceList1594076885316/lightreadingredhatopenshiftu2019sroadtothenetworkedgev3rg1594076883676.pdf

[163] 'ACX500, ACX1000, ACX2000, and ACX4000 Universal Metro Routers - Juniper Network', Juniper Networks. Accessed: Aug. 07, 2020. [Online]. Available: https://www.juniper.net/us/en/products-services/routing/acx-series/datasheets/1000397.page

[164] *Leveraging the Cloud in the New 5G Era (2538862)*, (Sep. 10, 2020). Accessed: Sep. 17, 2020. [Online Video]. Available: https://event.on24.com/eventRegistration/console/EventConsoleApollo.jsp?&eventid=2538862&sessionid=1&username=&partnerref=&format=fhaudio&mobile=&flashsupportedmobiledevice=&helpcenter=&key=B15CB82BEFB831C3B67EAECF0A69E487&newConsole=false&nxChe=true&text_language_id=en&playerwidth=748&playerheight=526&eventuserid=346459586&contenttype=A&mediametricsessionid=302821810&mediametricid=3574887&usercd=346459586&mode=launch

[165] 'Building the Road to the Network Edge: A Look at Container Innovation', Light Reading. Accessed: Aug. 20, 2020. [Online]. Available: https://www.lightreading.com/webinar.asp?webinar_id=1656&webinar_promo=5518&utm_source=light-reading&utm_medium=email&utm_campaign=NNRedHatWB

[166] 'New Packet Network Design for Transporting 5G Converged Traffic - Ciena', Ciena, White paper, 2019. Accessed: Aug. 20, 2020. [Online]. Available: http://www.ciena.com/insights/white-papers/new-packet-network-design-for-transporting-5g-converged-traffic.html

[167] 'Sharpening the Edge: Overview of the LF Edge Taxonomy and Framework', Linux Foundation, Jul. 2020. Accessed: Sep. 16, 2021. [Online]. Available: https://www.lfedge.org/wp-content/uploads/2020/07/LFedge_Whitepaper.pdf

[168] '5G; System architecture for the 5G System (5GS)'. ETSI TS 123 501, European Telecommunications Standards Institute, Sophia Antipolis, Jul. 2023. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/16.17.00_60/ts_123501v161700p.pdf

[169] '5G; NG-RAN; NG general aspects and principles'. ETSI TS 138 410, European Telecommunications Standards Institute, Sophia Antipolis, France, Oct. 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138400_138499/138410/16.04.00_60/ts_138410v160400p.pdf

[170] S. Redana *et al.*, 'View on 5G Architecture', 5G PPP, Feb. 2020. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2020/02/5G-PPP-5G-Architecture-White-Paper_final.pdf

[171] 'Study on new radio access technology: Radio access architecture and interfaces'. 3GPP TR 38.801, 3rd Generation Partnership Project, Sophia Antipolis, France, Apr. 03, 2017. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.801/38801-e00.zip

[172] 'Transport network support of IMT-2020/5G', ITU-T GSTR-TN5G, International Telecommunication Union, Oct. 2018. [Online]. Available: https://www.itu.int/hub/publication/t-tut-home-2018/

[173] 'Architecture of the metro transport network'. Rec. ITU-T G.8310, International Telecommunication Union, Dec. 2020. [Online]. Available: https://www.itu.int/rec/T-REC-G.8310/en

[174] 'Characteristics of transport networks to support IMT-2020/5G', Rec. ITU-T G.8300, International Telecommunication Union, May 2020. [Online]. Available: https://www.itu.int/rec/T-REC-G.8300/en

[175] 'Unified functional architecture of transport networks'. Rec. ITU-T G.800, International Telecommunication Union, Apr. 2016. [Online]. Available: https://www.itu.int/rec/T-REC-G.800/en

[176] 'Interfaces for metro transport networks'. Rec. ITU-T G.8312, International Telecommunication Union, Dec. 2020. [Online]. Available: https://www.itu.int/rec/T-REC-G.8312/en

[177] '5G wireless fronthaul requirements in a passive optical network context'. Rec. ITU-T G.Sup66, International Telecommunication Union, Jul. 2019. [Online]. Available: https://www.itu.int/rec/T-REC-G.Sup66/en

[178] 'Application of optical transport network recommendations to 5G transport'. Rec. ITU-T G.Sup67, International Telecommunication Union, Jul. 2019. [Online]. Available: https://www.itu.int/rec/T-REC-G.Sup67/en

[179] H. Nussbacher, 'Bottlenecks and link upgrades', Aug. 12, 2020. Accessed: Sep. 12, 2020. [Online]. Available: https://www.mail-archive.com/nanog@nanog.org/msg108720.html

[180] *Differentiated Services with Next-Generation Transport Network*, (Sep. 16, 2021). [Online Video]. Available: https://event.on24.com/wcc/r/3299440/0DBD07EBD87C741DC3A0B3663B0C948C/2669986

[181] D. Welch *et al.*, 'Point-to-Multipoint Optical Networks Using Coherent Digital Subcarriers', *J. Light. Technol.*, vol. 39, no. 16, pp. 5232–5247, Aug. 2021, doi: 10.1109/JLT.2021.3097163.

[182] *Metro Optical Architectures for the Future*, (May 25, 2021). Accessed: Nov. 01, 2021. [Online Video]. Available: https://event.on24.com/eventRegistration/EventLobbyServletV2?target=reg20V2.jsp&eventid=3072283&sessionid=1&key=5F19146276A1A99DC9434E7A3D81EB4F&groupId=2194889&partnerref=LReventspage&sourcepage=register

[183] 'Global Information Infrastructure Scenario Methodology'. Rec. ITU-T Y.120, International Telecommunication Union, Jun. 1998. [Online]. Available: https://www.itu.int/rec/T-REC-Y.120/en

[184] 'Optical transport network physical layer interfaces'. Rec. ITU-T G.959.1, International Telecommunication Union, Jul. 2018. [Online]. Available: https://www.itu.int/rec/T-REC-G.959.1/en

[185] 'Core Network Architecture Recommendations for Access to Legacy Data Networks over ADSL'. TR-025, Broadband Forum, Sep. 1999. [Online]. Available: https://www.broadband-forum.org/pdfs/tr-025-1-0-0.pdf

[186] 'Data-Over-Cable Service Interface Specifications; Modular Headend Architecture; EQAM Architectural Overview Technical Report'. CM-TR-MHA-V02-081209, CableLabs, Dec. 09, 2008. [Online]. Available: https://www.cablelabs.com/specifications/modular-headend-architecture-technical-report

[187] 'Gigabit-capable passive optical networks (GPON): General characteristics'. Rec. ITU-T G.984.1, International Telecommunication Union, Mar. 2008. [Online]. Available: https://www.itu.int/rec/T-REC-G.984.1/en

[188] S. McQuerry, *Interconnecting Cisco Networking Devices (ICND)*, 2nd ed. in CCNA Self-Study. 800 East 96th Street, 3rd Floor, Indianapolis, IN 46240, USA: Cisco Press.

[189] 'WLR3 Customer reference pack'. Openreach Limited. Accessed: Apr. 30, 2022. [Online]. Available: https://www.openreach.co.uk/orpg/home/products/wlr3/downloads/WLR3Trainingpackv5final.ppt

[190] 'Drop Passives: Splitters, Couplers and Power Inserters'. ANSI/SCTE 153 2021, Society of Cable Telecommunications Engineers, Exton, PA, USA, 2021.

[191] '40-Gigabit-capable passive optical networks (NG-PON2): General requirements'. Rec. ITU-T G.989.1, International Telecommunication Union, Mar. 2013. [Online]. Available: https://www.itu.int/rec/T-REC-G.989.1/en

[192] '10-Gigabit-capable passive optical networks (XG-PON): General requirements'. Rec. ITU-T G.987.1, International Telecommunication Union, Mar. 2016. [Online]. Available: https://www.itu.int/rec/T-REC-G.987.1/en

[193] J. Chamberlain and M. Alrutz, 'Maintaining the power advantage in HFC networks', presented at the SCTE-ISBA-NCTA 2017 Fall Technical Forum, Denver, Colorado, USA, Oct. 2017.

[194] 'Generic functional architecture of the optical media network'. Rec. ITU-T G.807, International Telecommunication Union, Feb. 2020. [Online]. Available: https://www.itu.int/rec/T-REC-G.807/en

[195] 'Universal Mobile Telecommunications System (UMTS); General Universal Mobile Telecommunications System (UMTS) architecture'. ETSI TS 123 101, European Telecommunications Standards Institute, Sophia Antipolis, France, Jul. 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123100_123199/123101/16.00.00_60/ts_123101v160000p.pdf

[196] I. Pepelnjak, 'SDN Tip: API != SDN', Feb. 17, 2023. [Online]. Available: https://mail.google.com/mail/u/0/#search/automation+SDN+pepelnjak/FMfcgzGrcjKRHNmSbkfTkBsqSjxHmgCQ

[197] 'Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV', ETSI GR NFV 003 V1.7.1, Jan. 2023. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/NFV/001_099/003/01.07.01_60/gr_NFV003v010701p.pdf

[198] 'Network Functions Virtualisation (NFV); Management and Orchestration', Group Report ETSI GR NFV-MAN 001 V1.2.1, Dec. 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/NFV-MAN/001_099/001/01.02.01_60/gr_NFV-MAN001v010201p.pdf

[199] 'Advanced Configuration and Power Interface (ACPI) Specification Release 6.5', UEFI Forum, Aug. 2022. [Online]. Available: https://uefi.org/sites/default/files/resources/ACPI_Spec_6_5_Aug29.pdf

[200] 'Environmental Engineering (EE); Green Abstraction Layer (GAL); Power management capabilities of the future energy telecommunication fixed network nodes', ETSI Standard ETSI ES 203 237 V1.1.1, Mar. 2014.

[201] T. Hudek, 'Overview of the Power Management Framework - Windows drivers'. Accessed: May 29, 2023. [Online]. Available: https://learn.microsoft.com/en-us/windows-hardware/drivers/kernel/overview-of-the-power-management-framework

[202] 'Environmental Engineering (EE); Green Abstraction Layer (GAL); Power management capabilities of the future energy telecommunication fixed network nodes; Enhanced Interface for power management in Network Function Virtualisation (NFV) environments'. ETSI ES 203 682 V1.1.1, European Telecommunications Standards Institute, Sophia Antipolis, France, Feb. 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_es/203600_203699/203682/01.01.01_60/es_203682v010101p.pdf

[203] Z. Shang and K. Wolter, 'Delay Evaluation of OpenFlow Network Based on Queueing Model'. arXiv, Aug. 23, 2016. doi: 10.48550/arXiv.1608.06491.

[204] I. Pepelnjak, *SDN and OpenFlow - The Hype and the Harsh Reality*. ipSpace.net, 2014. [Online]. Available: https://www.ipspace.net/SDN_and_OpenFlow

[205] P. J. Leach, R. Salz, and M. H. Mealling, 'A Universally Unique IDentifier (UUID) URN Namespace', Internet Engineering Task Force, Request for Comments RFC 4122, Jul. 2005. doi: 10.17487/RFC4122.

[206] D. Reforgiato *et al.*, 'Exporting data-plane energy-aware capabilities from network devices toward the control plane: The Green Abstraction Layer', in *2012 17th European Conference on Networks and Optical Communications*, Jun. 2012, pp. 1–6. doi: 10.1109/NOC.2012.6249924.

[207] 'Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Or-Vi reference point - Interface and Information Model Specification', Group Specification ETSI GS NFV-IFA 005 V3.1.1, Aug. 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/nfv-ifa/001_099/005/03.01.01_60/gs_nfv-ifa005v030101p.pdf

[208] 'Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Vi-Vnfm reference point - Interface and Information Model Specification', Group Specification ETSI GS NFV-IFA 006 V3.5.1, May 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/006/03.05.01_60/gs_NFV-IFA006v030501p.pdf

[209] 'Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Or-Vnfm reference point - Interface and Information Model Specification', Group Specification ETSI GS

NFV-IFA 007 V3.5.1, May 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/007/03.05.01_60/gs_NFV-IFA007v030501p.pdf

[210] 'Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Ve-Vnfm reference point - Interface and Information Model Specification', Group Specification ETSI GS NFV-IFA 008 V3.5.1, May 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/008/03.05.01_60/gs_NFV-IFA008v030501p.pdf

[211] 'Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; VNF Descriptor and Packaging Specification', Group Specification ETSI GS NFV-IFA 011 V4.2.1. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/011/04.02.01_60/gs_NFV-IFA011v040201p.pdf

[212] 'Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Os-Ma-Nfvo reference point - Interface and Information Model Specification', Group Specification ETSI GS NFV-IFA 013 V3.4.1, Jun. 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/013/03.04.01_60/gs_NFV-IFA013v030401p.pdf

[213] 'Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Network Service Templates Specification', Group Specification ETSI GS NFV-IFA 014 V4.2.1, May 2021. [Online]. Available: https://docbox.etsi.org/isg/nfv/open/Publications_pdf/Specs-Reports/NFV-IFA%20014v4.2.1%20-%20GS%20-%20Network%20Service%20Templates%20Spec.pdf

[214] 'Network Functions Virtualisation (NFV); Use Cases', Group Report ETSI GR NFV 001 V1.3.1, Mar. 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/NFV/001_099/001/01.03.01_60/gr_NFV001v010301p.pdf

[215] 'Windows Internals, Part 1: System architecture, processes, threads, memory management, and more, 7th Edition', Accessed: Jul. 27, 2020. [Online]. Available: /content/one-dot-com/one-dot-com/us/en/higher-education/program.html

[216] D. Wu and L. Cui, 'A comprehensive survey on Segment Routing Traffic Engineering', *Digit. Commun. Netw.*, Feb. 2022, doi: 10.1016/j.dcan.2022.02.006.

[217] C.-H. Lung and H. ElBakoury, 'Exploiting Segment Routing and SDN Features for Green Traffic Engineering', in *2022 IEEE 8th International Conference on Network Softwarization (NetSoft)*, Jun. 2022, pp. 49–54. doi: 10.1109/NetSoft54395.2022.9844091.

[218] C. Filsfils, S. Previdi, G. Dawra, and W. Henderickx, 'Interconnecting Millions of Endpoints with Segment Routing', Internet Engineering Task Force, Request for Comments RFC 8604, Jun. 2019. doi: 10.17487/RFC8604.

[219] A. Kaul, 'Spring for service provider networks', presented at the SANOG 34, Aug. 02, 2019. Accessed: Jun. 14, 2023. [Online]. Available: https://www.sanog.org/resources/sanog34/SANOG34-Tutorials-spring-for-service-providers-v1.3.pdf

[220] J. Moy, 'OSPF Version 2', Internet Engineering Task Force, Request for Comments RFC 2328, Apr. 1998. doi: 10.17487/RFC2328.

[221] 'Who We Are', CNCF. Accessed: Jun. 03, 2024. [Online]. Available: https://www.cncf.io/about/who-we-are/

[222] E.-V. Depasquale, F. Davoli, and H. Rajput, 'Dynamics of Research into Modeling the Power Consumption of Virtual Entities Used in the Telco Cloud', *Sensors*, vol. 23, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/s23010255.

[223] E.-V. Depasquale, H. A. Salam, and F. Davoli, 'PAD: A graphical and numerical enhancement of structural coding to facilitate thematic analysis of a literature corpus', *MethodsX*, vol. 9, p. 101633, Jan. 2022, doi: 10.1016/j.mex.2022.101633.

[224] 'DPDK', *DPDK*, 2021, Accessed: Feb. 25, 2021. [Online]. Available: https://www.dpdk.org/

[225] S. Fu, J. Liu, and W. Zhu, 'Multimedia Content Delivery with Network Function Virtualization: The Energy Perspective', *IEEE Multimed.*, vol. 24, no. 3, pp. 38–47, 2017, doi: 10.1109/MMUL.2017.3051514.

[226] Q. Chen, P. Grosso, K. van der Veldt, C. de Laat, R. Hofman, and H. Bal, 'Profiling Energy Consumption of VMs for Green Cloud Computing', in *2011 IEEE Ninth International Conference*

on Dependable, Autonomic and Secure Computing, Sydney, Australia: IEEE, Dec. 2011, pp. 768–775. doi: 10.1109/DASC.2011.131.

[227] N. Khan, 'Investigating Energy Efficiency of Physical and Virtual Machines in Cloud Computing', Univeristy of Oslo, 2017. Accessed: Nov. 07, 2019. [Online]. Available: https://www.duo.uio.no/handle/10852/61321

[228] G. Li, D. Zhang, Y. Li, and K. Li, 'Toward energy-efficiency optimization of pktgen-DPDK for green network testbeds', *China Commun.*, vol. 15, no. 11, pp. 199–207, Nov. 2018, doi: 10.1109/CC.2018.8543100.

[229] 'Intel® Ethernet Server Adapter I350: Product Brief', *Intel*, Mar. 2021, Accessed: Feb. 09, 2021. [Online]. Available: https://www.intel.com/content/www/us/en/products/docs/network-io/ethernet/10-25-40-gigabit-adapters/ethernet-i350-server-adapter-brief.html

[230] Z. Xu, F. Liu, T. Wang, and H. Xu, 'Demystifying the energy efficiency of Network Function Virtualization', in *2016 IEEE/ACM 24th International Symposium on Quality of Service (IWQoS)*, Jun. 2016, pp. 1–10. doi: 10.1109/IWQoS.2016.7590429.

[231] Z. Zhang and S. Fu, 'Macropower: A coarse-grain power profiling framework for energy-efficient cloud computing', in *30th IEEE International Performance Computing and Communications Conference*, Nov. 2011, pp. 1–8. doi: 10.1109/PCCC.2011.6108061.

[232] R. Morabito, 'Power Consumption of Virtualization Technologies: An Empirical Investigation', in *2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC)*, Dec. 2015, pp. 522–527. doi: 10.1109/UCC.2015.93.

[233] R. Shea, H. Wang, and J. Liu, 'Power consumption of virtual machines with network transactions: Measurement and improvements', in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, Apr. 2014, pp. 1051–1059. doi: 10.1109/INFOCOM.2014.6848035.

[234] M. Aldossary, K. Djemame, I. Alzamil, A. Kostopoulos, A. Dimakis, and E. Agiatzidou, 'Energy-aware cost prediction and pricing of virtual machines in cloud computing environments', *Future Gener. Comput. Syst.*, vol. 93, pp. 442–459, Apr. 2019, doi: 10.1016/j.future.2018.10.027.

[235] Y. Li, Y. Wang, B. Yin, and L. Guan, 'An Online Power Metering Model for Cloud Environment', in *2012 IEEE 11th International Symposium on Network Computing and Applications*, Cambridge, MA, Aug. 2012, pp. 175–180. doi: 10.1109/NCA.2012.10.

[236] C. Wen, X. Long, Y. Yang, F. Ni, and Y. Mu, 'System Power Model and Virtual Machine Power Metering for Cloud Computing Pricing', in *2013 Third International Conference on Intelligent System Design and Engineering Applications*, Jan. 2013, pp. 1379–1382. doi: 10.1109/ISDEA.2012.327.

[237] P. Arroba, J. L. Risco-Martín, M. Zapater, J. M. Moya, J. L. Ayala, and K. Olcoz, 'Server Power Modeling for Run-time Energy Optimization of Cloud Computing Facilities', *Energy Procedia*, vol. 62, pp. 401–410, Jan. 2014, doi: 10.1016/j.egypro.2014.12.402.

[238] W. Wu, W. Lin, and Z. Peng, 'An intelligent power consumption model for virtual machines under CPU-intensive workload in cloud environment', *Soft Comput.*, vol. 21, no. 19, pp. 5755–5764, Oct. 2017, doi: 10.1007/s00500-016-2154-6.

[239] G. Dhiman, K. Mihic, and T. Rosing, 'A system for online power prediction in virtualized environments using Gaussian mixture models', in *Proceedings of the 47th Design Automation Conference*, Anaheim, CA: ACM Press, Jun. 2010, p. 807. doi: 10.1145/1837274.1837478.

[240] H. A. Salam, F. Davoli, and A. Timm-Giel, 'Improving Prediction Accuracy for Power Consumption in Virtual Environments', in *2019 29th International Telecommunication Networks and Applications Conference (ITNAC)*, Nov. 2019, pp. 1–6. doi: 10.1109/ITNAC46935.2019.9077952.

[241] H. A. Salam, F. Davoli, A. Carrega, and A. Timm-Giel, 'Towards Prediction of Power Consumption of Virtual Machines for Varying Loads', in *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, Nov. 2018, pp. 1–6. doi: 10.1109/ATNAC.2018.8615319.

[242] M. Callau-Zori, L. Samoila, A.-C. Orgerie, and G. Pierre, 'An experiment-driven energy consumption model for virtual machine management systems', *Sustain. Comput. Inform. Syst.*, vol. 18, pp. 163–174, Jun. 2018, doi: 10.1016/j.suscom.2017.11.001.

[243] S. Chinprasertsuk and S. Gertphol, ‘Power model for virtual machine in cloud computing’, in *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, May 2014, pp. 140–145. doi: 10.1109/JCSSE.2014.6841857.

[244] T. Enokido and M. Takizawa, ‘The Extended Power Consumption Model to Perform Computation Type Application Processes on Virtual Machines’, in *2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, Fukuoka, Japan: IEEE, Jul. 2016, pp. 15–22. doi: 10.1109/CISIS.2016.82.

[245] T. Enokido and M. Takizawa, ‘Power Consumption Model of a Server to Perform Communication Type Application Processes on Virtual Machines’, in *2015 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA)*, Nov. 2015, pp. 275–282. doi: 10.1109/BWCCA.2015.67.

[246] B. Krishnan, H. Amur, A. Gavrilovska, and K. Schwan, ‘VM power metering: feasibility and challenges’, *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 38, no. 3, p. 56, Jan. 2011, doi: 10.1145/1925019.1925031.

[247] C. Isci and M. Martonosi, ‘Runtime power monitoring in high-end processors: methodology and empirical data’, in *Proceedings. 36th Annual IEEE/ACM International Symposium on Microarchitecture, 2003. MICRO-36.*, Dec. 2003, pp. 93–104. doi: 10.1109/MICRO.2003.1253186.

[248] T. Enokido and M. Takizawa, ‘Power Consumption and Computation Models of Virtual Machines to Perform Computation Type Application Processes’, in *2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems*, Jul. 2015, pp. 126–133. doi: 10.1109/CISIS.2015.18.

[249] M. Aldossary and K. Djemame, ‘Performance and Energy-based Cost Prediction of Virtual Machines Live Migration in Clouds’:, in *Proceedings of the 8th International Conference on Cloud Computing and Services Science*, Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 384–391. doi: 10.5220/0006682803840391.

[250] M. Aldossary and K. Djemame, ‘Energy-based Cost Model of Virtual Machines in a Cloud Environment’, in *2018 Fifth International Symposium on Innovation in Information and Communication Technology (ISIICT)*, Oct. 2018, pp. 1–8. doi: 10.1109/ISIICT.2018.8613288.

[251] I. Alzamil and K. Djemame, ‘Energy Prediction for Cloud Workload Patterns’, in *Economics of Grids, Clouds, Systems, and Services*, J. Á. Bañares, K. Tserpes, and J. Altmann, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 160–174. doi: 10.1007/978-3-319-61920-0_12.

[252] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya, ‘Virtual machine power metering and provisioning’, in *Proceedings of the 1st ACM symposium on Cloud computing - SoCC ’10*, Indianapolis, Indiana, USA: ACM Press, 2010, p. 39. doi: 10.1145/1807128.1807136.

[253] I. Waßmann, D. Versick, and D. Tavangarian, ‘Energy consumption estimation of virtual machines’, in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, in SAC ’13. Coimbra, Portugal: Association for Computing Machinery, Mar. 2013, pp. 1151–1156. doi: 10.1145/2480362.2480579.

[254] C. Gu, P. Shi, S. Shi, H. Huang, and X. Jia, ‘A Tree Regression Based Approach for VM Power Metering’, *IEEE Access*, vol. 3, pp. 1–1, Jan. 2015, doi: 10.1109/ACCESS.2015.2430276.

[255] W. Silva-de-Souza *et al.*, ‘Containergy—A Container-Based Energy and Performance Profiling Tool for Next Generation Workloads’, *Energies*, vol. 13, no. 9, p. 2162, May 2020, doi: 10.3390/en13092162.

[256] T. Veni and S. M. S. Bhanu, ‘Prediction Model for Virtual Machine Power Consumption in Cloud Environments’, *Procedia Comput. Sci.*, vol. 87, pp. 122–127, 2016, doi: 10.1016/j.procs.2016.05.137.

[257] ‘SPEC Benchmarks and Tools’. Accessed: Jul. 16, 2023. [Online]. Available: https://www.spec.org/benchmarks.html

[258] J. Phung, C. L. Young, and A. Y. Zomaya, ‘Application-Agnostic Power Monitoring in Virtualized Environments’, in *2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, May 2017, pp. 335–344. doi: 10.1109/CCGRID.2017.100.

[259] J. Phung, Y. C. Lee, and A. Y. Zomaya, 'Lightweight Power Monitoring Framework for Virtualized Computing Environments', *IEEE Trans. Comput.*, vol. 69, no. 1, pp. 14–25, Jan. 2020, doi: 10.1109/TC.2019.2936018.

[260] R. Bolla, R. Bruschi, F. Davoli, and J. F. Pajo, 'A Model-Based Approach Towards Real-Time Analytics in NFV Infrastructures', *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 529–541, Jun. 2020, doi: 10.1109/TGCN.2019.2961192.

[261] A. E. Husain Bohra and V. Chaudhary, 'VMeter: Power modelling for virtualized clouds', in *2010 IEEE International Symposium on Parallel Distributed Processing, Workshops and Phd Forum (IPDPSW)*, Apr. 2010, pp. 1–8. doi: 10.1109/IPDPSW.2010.5470907.

[262] L. M. Moreira Zorello, M. G. Torres Vieira, R. A. Girani Tejos, M. A. Torres Rojas, C. Meirosu, and T. C. Melo de Brito Carvalho, 'Improving Energy Efficiency in NFV Clouds with Machine Learning', in *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, Jul. 2018, pp. 710–717. doi: 10.1109/CLOUD.2018.00097.

[263] J. Stoess, C. Lang, and F. Bellosa, 'Energy Management for Hypervisor-Based Virtual Machines', in *USENIX Annual Technical Conference*, Santa Clara, CA, Jun. 2007.

[264] X. Li, W. Cheng, T. Zhang, F. Ren, and B. Yang, 'Towards Power Efficient High Performance Packet I/O', *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 4, pp. 981–996, Apr. 2020, doi: 10.1109/TPDS.2019.2957746.

[265] 'Virtio - KVM', 2020, Accessed: Aug. 27, 2020. [Online]. Available: https://www.linux-kvm.org/page/Virtio

[266] 'Poll Mode Driver — Data Plane Development Kit 20.11.0-rc0 documentation', Dec. 2014, [Online]. Available: https://doc.dpdk.org/guides/prog_guide/poll_mode_drv.html

[267] 'The Zeek Network Security Monitor', *Zeek*, Mar. 2020, Accessed: Oct. 02, 2020. [Online]. Available: https://zeek.org/

[268] 'Snort - Network Intrusion Detection & Prevention System', May 2015, Accessed: Oct. 02, 2020. [Online]. Available: https://www.snort.org/

[269] 'Open vSwitch', 2020, Accessed: Aug. 27, 2020. [Online]. Available: https://www.openvswitch.org/

[270] L. Rizzo and G. Lettieri, 'VALE, a switched ethernet for virtual machines', in *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, in CoNEXT '12. New York, NY, USA: Association for Computing Machinery, Dec. 2012, pp. 61–72. doi: 10.1145/2413176.2413185.

[271] R. Bruschi, F. Davoli, P. Lago, and J. F. Pajo, 'Joint Power Scaling of Processing Resources and Consolidation of Virtual Network Functions', in *5th IEEE International Conference on Cloud Networking (Cloudnet)*, Oct. 2016, pp. 70–75. doi: 10.1109/CloudNet.2016.20.

[272] W. Wu, W. Lin, L. He, G. Wu, and C. Hsu, 'A Power Consumption Model for Cloud Servers Based on Elman Neural Network', *IEEE Trans. Cloud Comput.*, pp. 1–1, 2019, doi: 10.1109/TCC.2019.2922379.

[273] R. Łukasik, 'Run-to-completion model in Data Plane processing - Nokia Wrocław', Dec. 2019, Accessed: Sep. 30, 2020. [Online]. Available: https://nokiawroclaw.pl/blog/post/run-to-completion-model-in-data-plane-processing/

[274] C. Kai, V. Blesson, K. Peter, and N. D. S, 'Power Modelling for Heterogeneous Cloud-Edge Data Centers', *Adv. Parallel Comput.*, vol. 32, pp. 804–813, 2018, doi: 10.3233/978-1-61499-843-3-804.

[275] K. N. Khan, M. Hirki, T. Niemi, J. K. Nurminen, and Z. Ou, 'RAPL in Action: Experiences in Using RAPL for Power Measurements', *ACM Trans. Model. Perform. Eval. Comput. Syst.*, vol. 3, no. 2, p. 9:1-9:26, Mar. 2018, doi: 10.1145/3177754.

[276] I. Molnar, 'Performance Counters for Linux, v8', 2009, Accessed: Jul. 23, 2020. [Online]. Available: https://lwn.net/Articles/336542/

[277] R. Bolla, R. Bruschi, A. Carrega, and F. Davoli, 'Green network technologies and the art of trading-off', in *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Apr. 2011, pp. 301–306. doi: 10.1109/INFCOMW.2011.5928827.

[278] 'TPC-W Homepage', 2021, Accessed: Feb. 11, 2021. [Online]. Available: http://www.tpc.org/tpcw/

[279] ‘Advanced Configuration and Power Interface (ACPI) Specification — ACPI Specification 6.4 documentation’, Jan. 2021, Accessed: Mar. 02, 2021. [Online]. Available: https://uefi.org/specs/ACPI/6.4/

[280] C. K. Keong, K. T. Wei, A. A. A. Ghani, and K. Y. Sharif, ‘Toward using software metrics as indicator to measure power consumption of mobile application: A case study’, Dec. 2015, pp. 172–177. doi: 10.1109/MySEC.2015.7475216.

[281] K. Li, ‘Optimal configuration of a multicore server processor for managing the power and performance tradeoff’, *J. Supercomput.*, vol. 61, no. 1, pp. 189–214, Jul. 2012, doi: 10.1007/s11227-011-0686-1.

[282] C. Mobius, W. Dargie, and A. Schill, ‘Power Consumption Estimation Models for Processors, Virtual Machines, and Servers’, *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 6, pp. 1600–1614, Jun. 2014, doi: 10.1109/TPDS.2013.183.

[283] ‘L.1310 : Energy efficiency metrics and measurement methods for telecommunication equipment’, Jul. 2017, Accessed: Jul. 27, 2020. [Online]. Available: https://www.itu.int/rec/T-REC-L.1310/en

[284] ‘Measurement method for energy efficiency of network functions virtualization’, Nov. 2018. Accessed: Jul. 23, 2020. [Online]. Available: https://www.itu.int/rec/T-REC-L.1361/en

[285] J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach*. Morgan Kaufmann, 2017.

[286] T. Inoue, A. Aikebaier, T. Enokido, and M. Takizawa, ‘Algorithms for Selecting Energy-Efficient Storage Servers in Storage and Computation Oriented Applications’, in *2012 IEEE 26th International Conference on Advanced Information Networking and Applications*, Mar. 2012, pp. 920–927. doi: 10.1109/AINA.2012.136.

[287] W. Al-Zubaedi and H. S. Al-Raweshidy, ‘A parameterized and optimized BBU pool virtualization power model for C-RAN architecture’, in *IEEE EUROCON 2017 -17th International Conference on Smart Technologies*, Jul. 2017, pp. 38–43. doi: 10.1109/EUROCON.2017.8011074.

[288] R. Mijumbi, J. Serrat, J.-L. Gorricho, and J. Rubio-Loyola, ‘On the Energy Efficiency Prospects of Network Function Virtualization’, *ArXiv151200215 Cs*, Dec. 2015, Accessed: Jul. 15, 2020. [Online]. Available: http://arxiv.org/abs/1512.00215

[289] T. Dlamini, ‘Softwarization in Future Mobile Networks and Energy Efficient Networks’, *Mob. Comput.*, Nov. 2019, doi: 10.5772/intechopen.89607.

[290] C. Dumitrescu, ‘Design Patterns for Packet Procesing Applications on Multi-core Intel Architecture’, Dec. 2008, Accessed: Sep. 29, 2020. [Online]. Available: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/ia-multicore-packet-processing-paper.pdf

[291] M. Hähnel, B. Döbel, M. Völp, and H. Härtig, ‘Measuring energy consumption for short code paths using RAPL’, *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 40, no. 3, pp. 13–17, Jan. 2012, doi: 10.1145/2425248.2425252.

[292] ‘The 5G Core is Vital to Deliver the Promise of 5G’, TelecomTV. Accessed: Feb. 17, 2021. [Online]. Available: https://www.telecomtv.com/content/intel-vsummit-5g-ran-5g-core/the-5g-core-is-vital-to-deliver-the-promise-of-5g-39164/

[293] R. Bhyrraju *et al.*, ‘VNF Descriptor and Packaging Specification’, ETSI, Sophia Antipolis Cedex - FRANCE, ETSI GS NFV-IFA 011, Sep. 2019. Accessed: Feb. 18, 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/011/02.07.01_60/gs_NFV-IFA011v020701p.pdf

[294] ‘The NANOG February 2021 Archive by thread’, *NANOG Mailing List Archive*. 2021. Accessed: Mar. 05, 2021. [Online]. Available: https://mailman.nanog.org/pipermail/nanog/2021-February/thread.html

[295] W. Dargie, ‘A Stochastic Model for Estimating the Power Consumption of a Processor’, *IEEE Trans. Comput.*, vol. 64, no. 5, pp. 1311–1322, May 2015, doi: 10.1109/TC.2014.2315629.

[296] P. Garraghan, Y. Al-Anii, J. Summers, H. Thompson, N. Kapur, and K. Djemame, ‘A Unified Model for Holistic Power Usage in Cloud Datacenter Servers’, in *2016 IEEE/ACM 9th International Conference on Utility and Cloud Computing (UCC)*, Dec. 2016, pp. 11–19.

[297] A. F. Monteiro and O. Loques, ‘Quantum Virtual Machine: A Scalable Model to Optimize Energy Savings and Resource Management’, in *2015 27th International Symposium on Computer*

Architecture and High Performance Computing (SBAC-PAD), Oct. 2015, pp. 194–201. doi: 10.1109/SBAC-PAD.2015.24.

[298] S. Sharma, V. Chang, U. S. Tim, J. L. H. Wong, and S. K. Gadia, ‘Cloud-based emerging services systems’, *Int. J. Inf. Manag.*, pp. 1–12, 2016.

[299] M. Zakarya and L. Gillam, ‘An Energy Aware Cost Recovery Approach for Virtual Machine Migration’, in *Economics of Grids, Clouds, Systems, and Services*, J. Á. Bañares, K. Tserpes, and J. Altmann, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 175–190. doi: 10.1007/978-3-319-61920-0_13.

[300] T. Zhao, J. Wu, S. Zhou, and Z. Niu, ‘Energy-delay tradeoffs of virtual base stations with a computational-resource-aware energy consumption model’, in *2014 IEEE International Conference on Communication Systems*, Nov. 2014, pp. 26–30. doi: 10.1109/ICCS.2014.7024759.

[301] A. Marletta, ‘CPU Usage Limiter for Linux’, 2012, Accessed: Jul. 27, 2020. [Online]. Available: http://cpulimit.sourceforge.net/

[302] J. Son, T. He, and R. Buyya, ‘CloudSimSDN-NFV: Modeling and simulation of network function virtualization and service function chaining in edge computing environments’, *Softw. Pract. Exp.*, vol. 49, no. 12, pp. 1748–1764, Dec. 2019, doi: 10.1002/spe.2755.

[303] A. Vasan, A. Sivasubramaniam, V. Shimpi, T. Sivabalan, and R. Subbiah, ‘Worth their watts? - an empirical study of datacenter servers’, in *HPCA - 16 2010 The Sixteenth International Symposium on High-Performance Computer Architecture*, Jan. 2010, pp. 1–10. doi: 10.1109/HPCA.2010.5463056.

[304] D. A. Slomon, P. Yosifovich, M. E. Russinovich, and A. Lonescu, *System architecture, processes, threads, memory management, and more*, 7th ed. in Windows internals ; Part 1. Pearson Education, 2017. Accessed: Jul. 27, 2020. [Online]. Available: https://www.oreilly.com/library/view/windows-internals-seventh/9780133986471/

[305] ‘CPU Analysis’, *Micrsoft*. May 2017. Accessed: Jul. 27, 2020. [Online]. Available: https://docs.microsoft.com/en-us/windows-hardware/test/wpt/cpu-analysis

[306] M. Goraczko, ‘Joulemeter: Computational Energy Measurement and Optimization’, *Microsoft Research*. Feb. 2010. Accessed: Jul. 27, 2020. [Online]. Available: https://www.microsoft.com/en-us/research/project/joulemeter-computational-energy-measurement-and-optimization/

[307] W. Lin, W. Wu, H. Wang, J. Z. Wang, and C.-H. Hsu, ‘Experimental and quantitative analysis of server power model for cloud data centers’, *Future Gener. Comput. Syst.*, vol. 86, pp. 940–950, Sep. 2018, doi: 10.1016/j.future.2016.11.034.

[308] M. Aldossary, I. Alzamil, and K. Djemame, ‘Towards Virtual Machine Energy-Aware Cost Prediction in Clouds’, in *Economics of Grids, Clouds, Systems, and Services*, C. Pham, J. Altmann, and J. Á. Bañares, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 119–131. doi: 10.1007/978-3-319-68066-8_10.

[309] ‘Intel® Core^{TM} i5-3230M Processor (3M Cache, up to 3.20 GHz) rPGA Product Specifications’, May 2014, Accessed: Jul. 27, 2020. [Online]. Available: https://ark.intel.com/content/www/us/en/ark/products/72164/intel-core-i5-3230m-processor-3m-cache-up-to-3-20-ghz-rpga.html

[310] ‘Mobile 3rd Generation Intel®CoreTM Processor Family, Mobile Intel® Pentium® Processor Family and Mobile Intel® Celeron® Processor Family’, Datasheet 326768–006, Jun. 2013. Accessed: Jul. 27, 2020. [Online]. Available: https://www.intel.com/content/dam/www/public/us/en/documents/datasheets/3rd-gen-core-family-mobile-vol-1-datasheet.pdf

[311] M. Doleželová *et al.*, ‘Power Management Guide Red Hat Enterprise Linux 6’, *Red Hat Customer Portal*. Dec. 2010. Accessed: Jul. 27, 2020. [Online]. Available: https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/6/html/power_management_guide/index

[312] M. Chios *et al.*, ‘Network Functions Virtualisation. An Introduction, Benefits, Enablers, Challenges & Call for Action’, in *SDN and OpenFlow World Congress*, Darmstadt, Germany, Oct. 2012.

[313] D. Molka, D. Hackenberg, R. Schöne, and M. S. Müller, ‘Characterizing the energy consumption of data transfers and arithmetic operations on x86-64 processors’, in *International Conference on Green Computing*, Aug. 2010, pp. 123–133. doi: 10.1109/GREENCOMP.2010.5598316.

[314] A. Rahman, P. Lynch, P. Aranda, C. Bernardos, J. Zuniga, and L. Contreras, ‘Network Virtualization Research Challenges’, Request For Comments (RFC) - Informational RFC 8568, ISSN: 2070-1721, Apr. 2019. Accessed: Jul. 10, 2020. [Online]. Available: https://tools.ietf.org/html/rfc8568\#page-18

[315] Dahyun, ‘SR-IOV for NFV solutions - practical considerations and thoughts’, Oct. 2017, Accessed: Jul. 27, 2020. [Online]. Available: http://man-glove.blogspot.com/2017/10/sr-iov-for-nfv-solutions-practical.html

[316] ‘Fifth Generation Fixed Network (F5G); F5G Network Architecture Release 2’, Group Specification ETSI GS F5G 014 V1.1.1, May 2023. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/F5G/001_099/013/01.01.01_60/gs_F5G013v010101p.pdf

[317] S. Shew, ‘Towards a journal paper’, Nov. 02, 2022. [Online]. Available: https://drive.google.com/open?id=1c6ha3HIoNwlRsJ4k7wy4qJo-pME9QWYd&usp=drive_fs

[318] E.-V. Depasquale, ‘Review of SGA + NOG Data’, Nov. 22, 2022. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/dc3a3d2b-e188-4ca3-bd37-a15b34921802/file_downloaded

[319] S. Perrin, ‘Validity on the use of questionnaires as a research tool’, Feb. 17, 2023. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/2e10f90e-38d4-47fe-9a2f-2fcebc9bf5a1/file_downloaded

[320] M. Tinka, ‘Would you help me craft a short questionnaire?’, Jun. 18, 2022. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/fc5adfbe-39a5-4f58-83ed-e4514fd94d14/file_downloaded

[321] S. Perrin, ‘Expressing an interest in collaboration’, Jun. 07, 2022. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/5878186b-234d-41ab-b69a-01a78ded7417/file_downloaded

[322] E.-V. Depasquale, ‘Respondents that present some inconsistency’, Sep. 20, 2022. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/d0317e21-2a22-480a-9299-5b166d3bab2a/file_downloaded

[323] P. Smith, ‘Review of SGA + NOG Data’, Mar. 20, 2023. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/a0a4bc7f-db2e-45c9-8920-c0fd056a5fe6/file_downloaded

[324] P. Arora, ‘Following this morning’s conversation - Number of respondents’, Jun. 08, 2022. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/a15feb1d-dad9-45d2-97be-c4e4260f246d/file_downloaded

[325] ‘Interfaces for the optical transport network’. Rec. ITU-T G.709/Y.1331, International Telecommunication Union, Jun. 06, 2020. [Online]. Available: https://www.itu.int/rec/T-REC-G.709-202006-I/en

[326] E.-V. Depasquale, ‘Routed optical networks’, May 01, 2023. Accessed: Jun. 21, 2023. [Online]. Available: https://mailman.nanog.org/pipermail/nanog/2023-May/222054.html

[327] *Ovidiu Roset review meeting recording*, (Apr. 27, 2023). [MP4]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/e3bafe8e-494a-4bcb-8b78-2d8db0525748/file_downloaded

[328] *Haider Khalid review meeting recording*, (May 22, 2023). [MP4]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/093db0a7-c69f-40eb-95ff-ae1cf54684ba/file_downloaded

[329] M. Bocci, T. Nadeau, L. Martini, C. Metz, and M. Aissaoui, ‘Segmented Pseudowire’, Internet Engineering Task Force, Request for Comments RFC 6073, Jan. 2011. doi: 10.17487/RFC6073.

[330] N. Leymann, B. Decraene, C. Filsfils, M. Konstantynowicz, and D. Steinberg, ‘Seamless MPLS Architecture’, Internet Engineering Task Force, Internet Draft draft-ietf-mpls-seamless-mpls-07, Jun. 2014. Accessed: Jul. 14, 2023. [Online]. Available: https://datatracker.ietf.org/doc/draft-ietf-mpls-seamless-mpls-07

[331] K. Haider, ‘Further assessment of survey results’, Jun. 2023. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/079d7b49-4232-4709-bd88-1e624c71df00/file_downloaded

[332] D. King, ‘The brief’, May 25, 2023. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/a99a03fc-0e18-40f9-a45f-3ca4bdcec931/file_downloaded

[333] *Dave Eilert review meeting recording*, (Apr. 21, 2023). [MP4]. Available: https://drive.google.com/open?id=1Gw0b0eegRFJyo7CwUqi4mEI0Spya8ZHq&usp=drive_fs

[334] M. Tinka, ‘Jupyter notebooks are ready’, Nov. 14, 2022. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/fba419c4-45c6-4332-917a-24babad0935b/file_downloaded

[335] ‘Multi-access Edge Computing (MEC); Framework and Reference Architecture’. ETSI GS MEC 003 V3.1.1, Sophia Antipolis, France, Mar. 2022. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/03.01.01_60/gs_MEC003v030101p.pdf

[336] ‘Metro Compute Networking: Use Cases and High Level Requirements’. TR-466 Issue 1, Broadband Forum, Sep. 2021.

[337] ‘Multi-service Broadband Network Architecture and Nodal Requirements’. TR-178 Issue 2, Broadband Forum, Sep. 2017. [Online]. Available: https://www.broadband-forum.org/technical/download/TR-178_Issue-2.pdf

[338] ‘Juniper Networks Metro Ethernet Design and Implementation Guide’. Juniper Networks, Inc., Aug. 2016. [Online]. Available: https://www.juniper.net/documentation/en_US/release-independent/solutions/information-products/pathway-pages/solutions/metro-ethernet-dg.pdf

[339] E. C. Rosen and Y. Rekhter, ‘Carrying Label Information in BGP-4’, Internet Engineering Task Force, Request for Comments RFC 3107, May 2001. doi: 10.17487/RFC3107.

[340] ‘IEEE Standard for Local and Metropolitan Area Network–Bridges and Bridged Networks’. IEEE Std 802.1Q-2018 (Revision of IEEE Std 802.1Q-2014), IEEE, Jul. 2018. doi: 10.1109/IEEESTD.2018.8403927.

[341] J. Drake *et al.*, ‘BGP MPLS-Based Ethernet VPN’, Internet Engineering Task Force, Request for Comments RFC 7432, Feb. 2015. doi: 10.17487/RFC7432.

[342] M. Schiano, A. Percelsi, and M. Quagliotti, ‘Flexible node architectures for metro networks’, in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2015, pp. 1–3. doi: 10.1364/OFC.2015.W3J.4.

[343] O. Gerstel *et al.*, ‘Multi-layer capacity planning for IP-optical networks’, *IEEE Commun. Mag.*, vol. 52, no. 1, pp. 44–51, Jan. 2014, doi: 10.1109/MCOM.2014.6710063.

[344] Anon., ‘Looking for insight’, Aug. 05, 2019. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/2efd65b4-8ae9-463d-9a89-e937d6ab2566/file_downloaded

[345] M. Tinka, ‘MPLS = Core’s data plane, IP = Core’s control plane?’, Feb. 12, 2023. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/d50c27b0-0491-4828-8459-ff6e7353c94a/file_downloaded

[346] ‘IEEE Standard for Local and Metropolitan Area Networks–Bridges and Bridged Networks’, *IEEE Std 8021Q-2022 Revis. IEEE Std 8021Q-2018*, pp. 1–2163, 2022, doi: 10.1109/IEEESTD.2022.10004498.

[347] ‘Ethernet ring protection switching’. Rec. ITU-T G.8032/Y.1344, International Telecommunication Union, Mar. 2020. [Online]. Available: https://www.itu.int/rec/T-REC-G.8032-202003-I/en

[348] ‘Ethernet linear protection switching’. Rec. ITU-T G.8031/Y.1342, International Telecommunication Union, Jan. 2015. [Online]. Available: https://www.itu.int/rec/T-REC-G.8031-201501-I/en

[349] M. Nurujjaman, S. Sebbah, and C. Assi, ‘Planning and optimization of highly resilient multi-ring ERP-based mesh networks’, in *2013 9th International Conference on the Design of Reliable Communication Networks (DRCN)*, Mar. 2013, pp. 203–209.

[350] W. Chen, Q. Yu, P. F. Yu, J. Ren, and Z. Huang, ‘Research of Ethernet Ring Protection Technology’, *Appl. Mech. Mater.*, vol. 556–562, pp. 6026–6029, May 2014, doi: 10.4028/www.scientific.net/AMM.556-562.6026.

[351] G. Tim, ‘Pseudowire Headend Termination (PWHT) For Juniper MX’, PACKETS AND STUFF. Accessed: Jul. 14, 2023. [Online]. Available: https://tgregory.org/2016/07/10/pseudowire-headend-termination-pwht-for-juniper-mx/

[352] B. Martin, ‘The future of transport in the metro area’, Aug. 01, 2019. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/4ef6849e-1c2c-415e-b142-1536fcb81f16/file_downloaded

[353] D. Lundén, ‘Technical question’, Aug. 07, 2019. [Online]. Available: https://data.mendeley.com/public-files/datasets/k72dntnfgv/files/2b3843c6-36c7-4637-8da0-2eeac64ccdd9/file_downloaded

[354] ‘External Network Network Interfaces (ENNI) and Operator Service Attributes’. Technical Specification MEF 26.2, The MEF Forum, Aug. 2016. [Online]. Available: https://wiki.mef.net/display/CESG/MEF+26.2+-+ENNI+and+Operator+Service+Attributes?preview=/59840549/60261919/MEF_26.2.pdf

[355] E. C. Rosen and L. Andersson, ‘Framework for Layer 2 Virtual Private Networks (L2VPNs)’, Internet Engineering Task Force, Request for Comments RFC 4664, Sep. 2006. doi: 10.17487/RFC4664.

[356] ‘5G; NR; NR and NG-RAN Overall description; Stage-2’. ETSI TS 138 300 V17.0.0, European Telecommunications Standards Institute, Sophia Antipolis, France, May 2022. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138300_138399/138300/17.00.00_60/ts_138300v170000p.pdf

[357] C. Castro *et al.*, ‘Scalable filterless coherent point-to-multipoint metro network architecture’, *J. Opt. Commun. Netw.*, vol. 15, no. 5, pp. B53–B66, May 2023, doi: 10.1364/JOCN.483130.

[358] M. M. Hosseini, J. Pedro, A. Napoli, N. Costa, J. E. Prilepsky, and S. K. Turitsyn, ‘Multi-period planning in metro-aggregation networks exploiting point-to-multipoint coherent transceivers’, *J. Opt. Commun. Netw.*, vol. 15, no. 3, pp. 155–162, Mar. 2023, doi: 10.1364/JOCN.475902.

[359] P. Pavon-Marino, N. Skorin-Kapov, and A. Napoli, ‘Tree-determination, routing, and spectrum assignment using point-to-multipoint coherent optics’, *J. Opt. Commun. Netw.*, vol. 15, no. 7, pp. C29–C40, Jul. 2023, doi: 10.1364/JOCN.482457.

[360] A. Teker, A. H. Örnek, and B. Canberk, ‘Network Bandwidth Usage Forecast in Content Delivery Networks’, in *2020 International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom)*, Jul. 2020, pp. 1–6. doi: 10.1109/CoBCom49975.2020.9174180.

[361] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, ‘FemtoCaching: Wireless Content Delivery Through Distributed Caching Helpers’, *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013, doi: 10.1109/TIT.2013.2281606.

[362] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, ‘Cache in the air: exploiting content caching and delivery techniques for 5G systems’, *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014, doi: 10.1109/MCOM.2014.6736753.

[363] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, ‘A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications’, *IEEE Access*, vol. 5, pp. 6757–6779, 2017, doi: 10.1109/ACCESS.2017.2685434.

[364] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, ‘A Comprehensive Survey on Fog Computing: State-of-the-Art and Research Challenges’, *IEEE Commun. Surv. Tutor.*, vol. 20, no. 1, pp. 416–464, 2018, doi: 10.1109/COMST.2017.2771153.

[365] J. Zhao, P. Liang, W. Liufu, and Z. Fan, ‘Recent Developments in Content Delivery Network: A Survey’, in *Parallel Architectures, Algorithms and Programming*, H. Shen and Y. Sang, Eds., Singapore: Springer, 2020, pp. 98–106. doi: 10.1007/978-981-15-2767-8_9.

[366] B. Zolfaghari *et al.*, ‘Content Delivery Networks: State of the Art, Trends, and Future Roadmap’, *ACM Comput. Surv.*, vol. 53, no. 2, p. 34:1-34:34, Apr. 2020, doi: 10.1145/3380613.

[367] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, ‘Energy Efficiency in the Future Internet: A Survey of Existing Approaches and Trends in Energy-Aware Fixed Network Infrastructures’, *IEEE Commun. Surv. Tutor.*, vol. 13, no. 2, pp. 223–244, 2011, doi: 10.1109/SURV.2011.071410.00073.

[368] A. K. Yadav, M. L. Garg, and Ritika, ‘Docker Containers Versus Virtual Machine-Based Virtualization’, in *Emerging Technologies in Data Mining and Information Security*, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, Eds., Singapore: Springer, 2019, pp. 141–150. doi: 10.1007/978-981-13-1501-5_12.

[369] S. Rivoire, P. Ranganathan, and C. Kozyrakis, ‘A comparison of high-level full-system power models’, in *Proceedings of the 2008 conference on Power aware computing and systems*, in HotPower’08. USA: USENIX Association, Dec. 2008, p. 3.

[370] K. Hinton, F. Jalali, and A. Matin, ‘Energy consumption modelling of optical networks’, *Photonic Netw. Commun.*, vol. 30, no. 1, pp. 4–16, Aug. 2015, doi: 10.1007/s11107-015-0491-5.

[371] W. Lin, H. Wang, and W. Wu, ‘A power monitoring system based on a multi-component power model’, *Int. J. Grid High Perform. Comput.*, vol. 10, no. 1, pp. 16–30, Jan. 2018.

[372] B. Rountree, D. K. Lowenthal, M. Schulz, and B. R. de Supinski, ‘Practical performance prediction under Dynamic Voltage Frequency Scaling’, in *2011 International Green Computing Conference and Workshops*, Jul. 2011, pp. 1–8. doi: 10.1109/IGCC.2011.6008553.

[373] T. Guérout, Y. Gaoua, C. Artigues, G. Da Costa, P. Lopez, and T. Monteil, ‘Mixed integer linear programming for quality of service optimization in Clouds’, *Future Gener. Comput. Syst.*, vol. 71, pp. 1–17, Jun. 2017, doi: 10.1016/j.future.2016.12.034.

[374] R. Bolla, R. Bruschi, F. Davoli, C. Lombardo, and N. S. Martinelli, ‘Analyzing the Power Consumption in Cloud-Native 5/6G Ecosystems’, in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2023, pp. 611–617. doi: 10.1109/ICCWorkshops57953.2023.10283755.

[375] ‘Environmental Engineering (EE); Green Abstraction Layer (GAL); Power management capabilities of the future energy telecommunication fixed network nodes; Enhanced Interface for power management in Network Functions Virtualisation (NFV) environments’. ETSI ES 203 682 V1.2.0, European Telecommunications Standards Institute, Sophia Antipolis, France, Jan. 2024.

[376] E. V. Depasquale, M. Tinka, S. Zammit, and F. Davoli, ‘A survey of trends and motivations regarding Communication Service Providers’ metro area network implementations’. Dec. 2023. doi: 10.13140/RG.2.2.13543.42400.

[377] ‘Powertop - ArchWiki’. Accessed: Feb. 20, 2024. [Online]. Available: https://wiki.archlinux.org/title/powertop

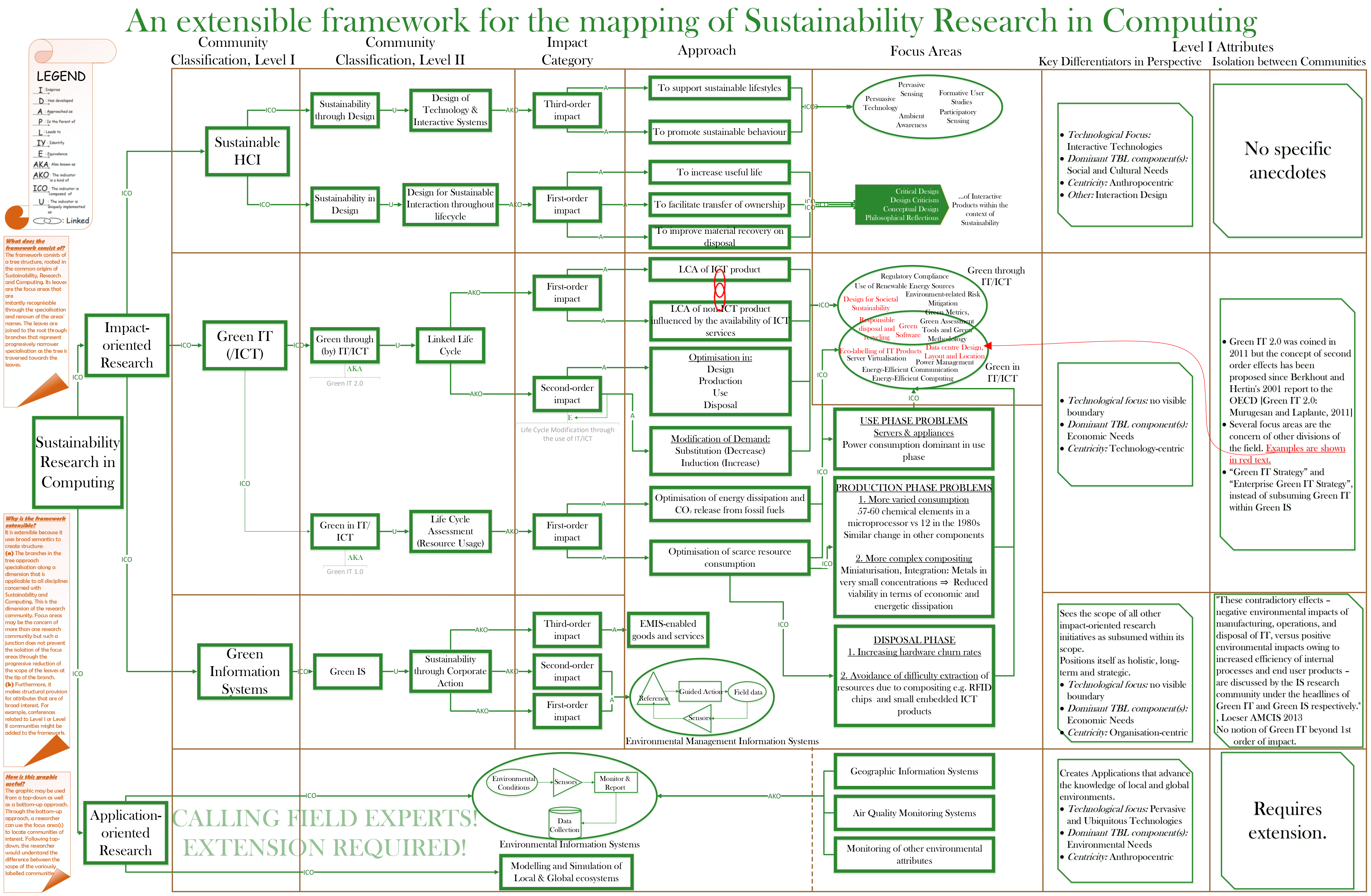
[378] A. van de Ven, ‘PowerTOP running average interval and display update interval’, Apr. 08, 2024.

[379] A. van de Ven, ‘PowerTOP running average interval and display update interval (update)’, Apr. 08, 2024.

[380] A. van de Ven, ‘PowerTOP running average interval and display update interval (2nd update)’, Apr. 30, 2024.

[381] J. Sultana, ‘Power measurement - Gen9’, May 24, 2024.

Appendix 1. An extensible framework for the mapping of Sustainability Research in Computing



An extensible framework for the mapping of Sustainability Research in Computing

What does the framework consist of?
The framework consists of a tree structure, rooted in the common origins of Sustainability, Research and Computing. Its leaves are the focus areas that are instantly recognisable through the specialisation and renown of the areas' names. The leaves are joined to the root through branches that represent progressively narrower specialisation as the tree is traversed towards the leaves.

Why is the framework extensible?
It is extensible because it uses broad semantics to create structure:
(a) The branches in the tree approach specialisation along a dimension that is applicable to all disciplines concerned with Sustainability and Computing. This is the dimension of the research community. Focus areas may be the concern of more than one research community but such a junction does not prevent the isolation of the focus areas through the progressive reduction of the scope of the leaves at the tip of the branch.
(b) Furthermore, it makes structural provision for attributes that are of broad interest. For example, conferences related to Level I or Level II communities might be added to the framework.

How is this graphic useful?
The graphic may be used from a top-down as well as a bottom-up approach. Through the bottom-up approach, a researcher can use the focus area(s) to locate communities of interest. Following top-down, the researcher would understand the difference between the scope of the variously labelled communities.

- *Technological Focus:* Interactive Technologies
- *Dominant TBL component(s):* Social and Cultural Needs
- *Centricity:* Anthropocentric
- *Other:* Interaction Design

No specific anecdotes

- *Technological focus:* no visible boundary
- *Dominant TBL component(s):* Economic Needs
- *Centricity:* Technology-centric

- Green IT 2.0 was coined in 2011 but the concept of second order effects has been proposed since Berkhout and Hertin's 2001 report to the OECD [Green IT 2.0: Murugesan and Laplante, 2011]
- Several focus areas are the concern of other divisions of the field. Examples are shown in red text.
- "Green IT Strategy" and "Enterprise Green IT Strategy", instead of subsuming Green IT within Green IS

Sees the scope of all other impact-oriented research initiatives as subsumed within its scope.
Positions itself as holistic, long-term and strategic.
- *Technological focus:* no visible boundary
- *Dominant TBL component(s):* Economic Needs
- *Centricity:* Organisation-centric

"These contradictory effects – negative environmental impacts of manufacturing, operations, and disposal of IT, versus positive environmental impacts owing to increased efficiency of internal processes and end user products – are discussed by the IS research community under the headlines of Green IT and Green IS respectively." , Loeser AMCIS 2013
No notion of Green IT beyond 1st order of impact.

Creates Applications that advance the knowledge of local and global environments.
- *Technological focus:* Pervasive and Ubiquitous Technologies
- *Dominant TBL component(s):* Environmental Needs
- *Centricity:* Anthropocentric

Requires extension.

Appendix 2. The PAD method for surveying a field of research.

This appendix regards the method used to survey the field of models of power consumption used in the telco cloud. It may be found online[71]. The novelty in the method is rooted in its objective of extracting researchers' *modus operandi*, instead of focusing solely on their research's output.

---

[71] https://www.sciencedirect.com/science/article/pii/S2215016122000188

Appendix 3. State of the Art in Video Distribution Networks

1. Overview

A researcher intent on improving the energy efficiency of a system is faced with a quest for intimate knowledge of the internals of the system under study. This requires an interest in the details of the organization of the system as well as its architecture. If the organization is concealed by the system's maker, the researcher is dependent upon abstractions that provide the required methods and data structures that reveal information about energy use. Makers with commercial interests (e.g. hardware manufacturers and closed-source software developers) characteristically shy away from revealing organizational details and therefore close collaboration at the abstraction layer is necessary. In this section, we explore detail about the system under study: the video distribution network. In particular, we are interested in its data plane functions, since this, as is generally true, is where most energy use takes place during operations.

The term "video distribution network" (VDN) encompasses a wide variety of system-architectures and technologies that ingest, process, aggregate and distribute video, audio and text information. The VDN is bounded at one end by the ingest to the headend, which egests to the transport network, and at the other end by the (VDN-dependent) network device that transmits the multimedia to the user terminal. Our scope excludes the architecture and organization of the television studio, although we do refer to it as a possible local source of multimedia when we refer to contribution as a means of local acquisition. Complementing this, we are concerned with the distribution of the product of the television studio. More generally, we are concerned with the distribution of the product of the television network, which distributes its product to the video distribution network over satellite, terrestrial and cable links.

We first introduce the genres of VDN. We refer to genres to classify identifiably different combinations of architectural and technological ingredients. Genres, therefore, differ not only in their architectures but also in the technologies out of which their organizations are assembled. The organization of the VDN's headend is notably genre-dependent and we dwell on the organizations from the perspective of their role in the delivery of the video service.

We proceed to suggest a template for the combination of these ingredients. The purpose of the suggested template is to illustrate a simplified representative implementation of the genre in terms of system-architecture and technology. We shall briefly indicate variants of this template. Variants have the major architectural features of the VDN-template but differ from one another in terms of macroscopic sequences (macro sequences, after N. Dang [138]). The hierarchy we shall use to organize the space of VDNs therefore consists of that shown in Fig. 9.

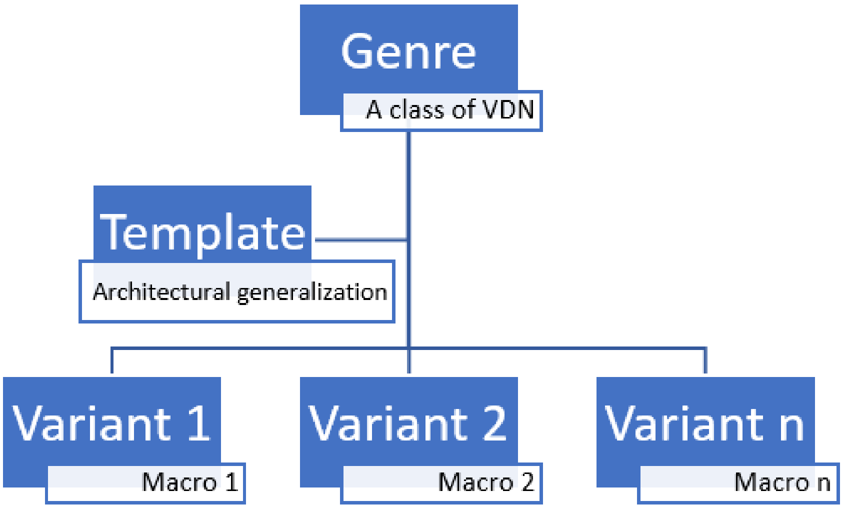


Fig. 9: Organisation of the VDN-architecture space

A VDN is owned either in full or in part by a Video Service Provider (VSP). Here, the term VSP is used to encompass any organization that delivers video to customers, regardless of the VDN's genre. We will then proceed to anecdotal references of what appears to be the VSP's ultimate destination, namely, the unified headend.

2. The genres

2.1 RFTV and IPTV. The first two genres we distinguish are RFTV and IPTV. RFTV distributes digital television in the fundamental organizational unit of the MPEG Transport Stream (MPEG-TS): it is the common denominator in all RFTV forms. An early moniker of IPTV was "telco TV" (ca. 2004), emphasizing its origin and its proponent as well as shedding some light on a commercial driver: opening up competition for "telco" operators with cable TV system operators.

RFTV minimally comprises formatting, source encoding, channel encoding modulation and a channel, all of which are familiar from digital communications theory. Formatting, source encoding, channel encoding and modulation (as well as other optional functions) are carried out at the headend (strictly, both source and channel encoding are optional but in practice they are not), where the video (and audio and text, where applicable) information enters the VDN for live broadcast transmission. Live broadcast transmission is specifically singled out as it requires a shorter introduction but Video on Demand (VoD) – a unicast form of transmission – can also be delivered from an RFTV headend. Other means of transmission, such as over-the-top (OTT), are better characterised as part of later generations of headend.

Similarities. IPTV and RFTV are similar in the following ways:

1. ***Common functions.*** IPTV transmission uses formatting and source encoding, like RFTV.
2. ***Managed networks.*** An important common characteristic is that both RFTV and IPTV use ***managed networks*** that are capable of guaranteeing QoS fit for ***tele***-vision.

Within this latter commonality, expressions of QoS in RFTV and IPTV are separated by the fundamental difference that IPTV packetizes the bit stream. Therefore, IPTV QoS can be expressed in terms of packet loss rate, packet latency and maximum variation in packet loss rate (jitter). The QoS required to support streaming video is well known. Examples are shown in Table 2 [140] and Table 3 [141]. "Max. setup time" is only relevant to connection-oriented network services (e.g. SDH/SONET connections). Here, a "network service" refers to an NP's product offering. It is the means through which the "user application" is transported from source to destination. Tolerable latency is large; similarly, ITU-T Standard G.1010 [141] places streaming video in the delay-tolerant class (~10 s) named "timely". This does seem rather unrealistic for viewing of live events, particularly those which have key moments (say, sport events) but is otherwise intuitively correct and this latency datum is repeated elsewhere [160]. Table 3 shows G.1010's specification of QoS for streaming video.

Table 2: Quality of service parameters [140]

User application	Max latency (s)	Max jitter (ms)	Pkt. loss (L3) (%)	Max. setup time	Min. avail (%)
Video on demand	2-20	50	0.5	seconds	99.5
Video broadcast	2-20	50	0.5	seconds	99.5

Table 3: Quality of service parameters [141]

User application	Latency (s)	Pkt. loss (%)
One-way video	< 10	< 1

Differences. IPTV and RFTV are fundamentally differentiated in the following ways:

1. ***Complete abstraction between source bit stream and channel.*** IPTV transmission uses formatting and source encoding, like RFTV, but between the source bit stream and the channel there is a complete abstraction. The rich diversity of modules that populate the layers between the application and the link (which includes the channel - item 5, above) of an Internet Protocol Suite architecture [139] may fill in the implementation. This is the current state of evolution of digital communication, where the Internet Protocol Suite architecture has provided the openness necessary for independent development of competing (same-layer) and complementary (dissimilar layer) modules.

2. ***Path selection***. IP is predicated upon the individually-routable packet and does not intrinsically require reservation of path resources like circuit-based communication. Evidently, this does not modify the characteristics of video's demands on its transport but it does allow transport to be provided more efficiently and possibly more flexibly than circuit-based transport. RFTV uses an isochronous MPEG-TS that supports no further path selection. Path selection is not intrinsically supported by MPEG-TS and therefore each MPEG-TS packet in RFTV follows the same physical path as its peers. RFTV is ***circuit-like***.
3. ***Multicast vs broadcast***. With IPTV, video is egested from the VDN headend as multicast over UDP for live transmissions and/or stored to be used for VoD. When stored video is requested by end-users that have passed authentication, authorization and accounting (AAA) management and digital rights management (DRM), the video is unicast using RTP over UDP. We have referred earlier to work that suggests that VoD is most responsible for growth of IP traffic in the metro core [97]. With RFTV, video is egested from the VDN headend as a channel broadcast.

2.2 Internet TV. A third genre is Internet TV, also referred to as OTT TV, or plainly OTT. The principal differentiator between Internet TV and the preceding two genres regards the network intermediate to the source and the destination. OTT is delivered over an intermediate network that is owned by an arbitrary combination of third parties whereas (as stated earlier) RFTV and IPTV are delivered over a managed network. Referring to Fig. 10 [145], the managed network may be owned by:

1. the Video Service Provider: as in cases (a),(b),(c) as well as case (d), the latter being that in which the VSP owns the active equipment
2. a Network Provider which leases a VPN to the VSP

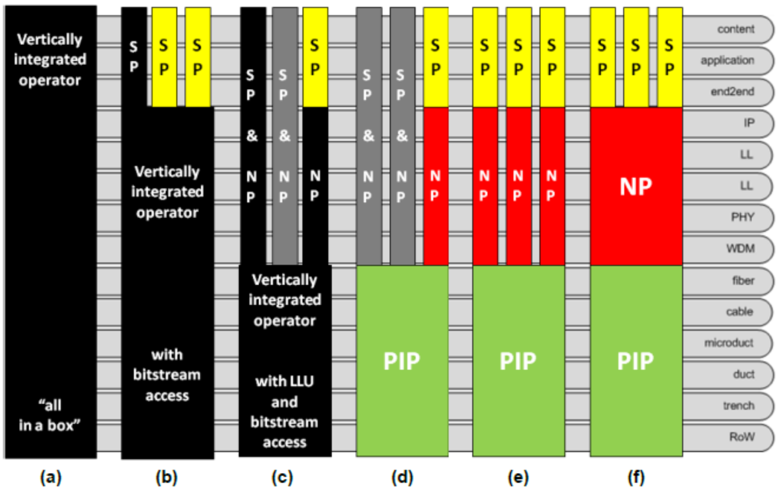


Fig. 10: Some of the business models enabled by layered networks [145]

This latter reference to VPN has purposely avoided specification of OSI layer, to conform to the broad characterization given in ITU-T Y.1311 [146]. Unlike Fig. 2(e) and Fig. 2(f), which relate to an L3 VPN, it is possible for a VSP to purchase L1 VPNs (e.g. OTNs) and L2 VPNs (e.g. Virtual Private LAN Service). These would form part of inter-metro distribution of the video signal. Therefore, there are even more business models than those suggested by Fig. 2, to reason about a VSP's position in the telecommunications layers.

The Content Delivery Network (CDN) is an important architectural component of this genre. One taxonomy [143] associates OTT delivery with a specific CDN architecture, where the CDN operator (e.g. Akamai or Limelight) owns the datacenters but not the network nor the content. A commonality is immediately visible: neither the content owner (aka: the rights holder and aka: the contributor), nor the VSP, nor the "content delivery accelerator" (the OTT CDN) own the intermediate network. This characteristic is pervasive in services delivered over-the-top. While this is not best-effort network service, for reasons that we shall enunciate, there is nonetheless a separation between the organization providing the service (the video) and that providing the network service. An evocative phrase often used to distinguish Internet TV from RFTV and IPTV is "walled garden". RFTV and IPTV are delivered in a walled garden, where QoS is within the direct control of the VSP. Internet TV is delivered over networks that are outside the direct control of the VSP.

Another observation concerns the range of content accessible over-the-top. Television content may be divided into (a) real-time, or linear in this industry's jargon and (b) stored content, which complements linear – i.e. anything that is not linear. Stored content is a good candidate for delivery through an OTT-CDN. Viewing of stored content, whether through IPTV (where it is commonly referred to as VoD) or OTT, covers a broad range of high-value content. Stored content includes content which was delivered linearly and recorded for delayed viewing. Delayed viewing is a growing behaviour and with at least one major North American television network, was reported as "the new normal" [148].

Linear content includes some high-value content in television, like sports, news and live events. It seems reasonable to add series premieres to this list, despite the (at least localized) pre-eminence of delayed viewing. It is evidently desirable to include linear content in an OTT-VSP's product range but technical challenges face aspiring adopters. Linear TV QoE and reliability are difficult problems to solve for an OTT-VSP without the combination of two stages of the delivery network: the Internet core and the regional network. There are at least two broad approaches towards achieving this goal:

- Customized cascading:
    - contracts with Tier-1 ISPs to obtain the reach of a global network. An example [149] illustrates the specialization of a Tier-1 ISP's network for video content distribution;
    - contracts with regional ISPs. An example [151] of such a contract includes the use of the regional ISPs' access network. The OTT reaches the region through the cloud.
- Turnkey cascading: Multi-CDN
    - contracts with multiple CDN operators [152] to ensure good global geographical coverage addresses both QoE and reliability concerns.

With these two conditions, the differentiator "managed network" must be replaced by a more focused differentiator to define a clear surface of demarcation between OTT and IPTV. The better differentiation is that described in Cisco's VNI FAQ [152], in distinguishing between "managed IP traffic" and "Internet traffic". Managed IP traffic is described as that which (a) flows only through a single network and (b) is managed by a single service provider. On the other hand, Internet traffic flows across an Internet backbone, crossing network boundaries between different (Internet) service providers and CDN providers, which therefore means that there are several, un-coordinated control planes.

Indeed, the geographical boundary of a VSP's delivery of linear content is likely to be the same as that of its managed network, whether physical, virtual or some turnkey wide-area networking solution that ingests the content and distributes it. This statement is a consolidation of the geographical scope of delivery of linear content and is valid for IPTV-VSPs as well as OTT-VSPs. The IPTV-VSP typically owns the network and is classifiable as a Network Provider whereas the OTT-VSP typically owns a part of the network. The IPTV-VSP typically is an incumbent in the telecommunications operators' field whereas the OTT-VSP typically is not. The OTT-VSP typically is an incumbent in the field of acquisition of content rights. The OTT-VSP typically must enter into contractual relationships (viz., the two approaches identified above) to ensure QoE and reliability.

A consequence of the lower end-to-end predictability of OTT's intermediate network is the development of a variety of application-layer protocols that support adaptive bit rate (ABR) streaming over HTTP. Examples include: Adobe's HTTP Dynamic Streaming (HDS), Apple's HTTP Live Streaming (HLS), Microsoft's HTTP Smooth Streaming (HSS) and ISO/IEC 23009-1 (better known as MPEG-DASH).

3. RFTV VDN Templates and Variants

The RFTV VDN template comprises (a) the RFTV headend and (b) a broadcast channel. The generalized, technology-agnostic template is represented by the well-known block diagram of a digital communication system and is reproduced in several textbooks [154, 155] on digital communications. The interested reader is referred to these sources for a system-level view of the technology-agnostic template.

The first specialization of the generalized template that leads towards the RFTV template is the MPEG-TS. The MPEG-TS is not part of the generalized form since its use in practice is specific to multimedia sources. Fig. 11 illustrates the relationship between the MPEG-TS and the source bits in a digital communications system.

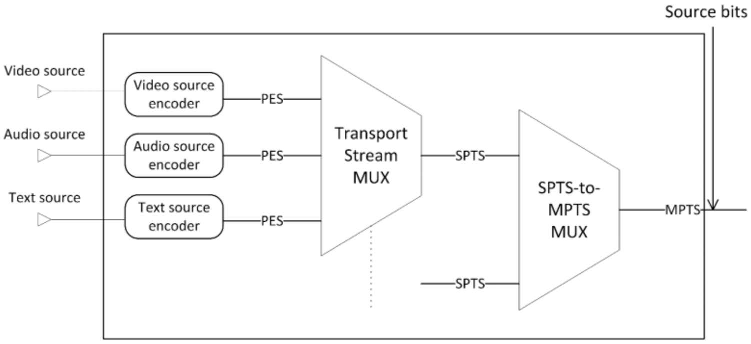


Fig. 11: The MPEG-TS is the source bit stream input to an RFTV system

The RFTV VDN template comprises a long-reach transport portion. The content source or the content aggregator uses satellite uplink to distribute to affiliates' headends. The affiliate uses a short-reach transport portion. Further differentiation within the short-reach transport leads directly to the variants. This follows because the principal differentiator from the agnostic form is the broadcast RF channel. Standardisation of RFTV system binds the channel to the carrier modulation and electrical characteristics of the transmitter and the receiver. For example, a cable TV transmitter in Europe modulates its carrier using 16-, 32-, 64-, 128- or 256-QAM; in North America, either 64-QAM or 256-QAM is selected by the transmitter. This statement is definitive because the variant (RFTV over cable) is standardised. This standardisation of the entire RFTV system reduces the task of mapping the division of the RFTV space into that of identifying the pertinent standards. Relevant system-level standard views are referred to in the following examples. The interested reader is referred to the indicated standard sources for a system-level view of the variants.

- Cable (e.g. CATV): RFTV distribution by cable networks [156]
    - Annex B representing US digital cable TV
    - Annex A, equivalent to ETSI EN 300 429 [167], represents European digital cable TV.
- Satellite: Digital Video Broadcasting – Satellite [157]
- Terrestrial: Digital Video Broadcasting – Terrestrial [158] and Advanced Television Steering Committee [159]

Despite the expectations that the OTT genre own the largest percentage of market share in the near future, at least in developed markets like the North American, European and the Far Eastern, a recent survey [169] is giving reason to re-evaluate conventional wisdom on future scenarios of genre adoption. It was found that in the five years between 2012 and 2017, homes in the USA that receive TV only through OTA broadcasts increased by 41% to 15.8 million. A particularly striking note in the press release is "greater access to … affordable entertainment" (bold italic font style added). It seems that price sensitivity is high in customers' ultimate selection of a genre for video service.

4. IPTV VDN Template and Variants

4.1 Template

IPTV implementations are diverse as a result of the decoupling between IP and the end-to-end path between the headend and the end user. Our effort here will be directed towards rationalising the diversity into a set of recognizable variants of the template.

Generalization is tractable, if we are willing to commit some depth of abstraction of the end-to-end path. This path may be coarsely divided into the long-haul IP transport infrastructure and metropolitan area transport infrastructure respectively. In this basic segmentation, the access segment lies within the metro-area segment. The junction of the long-haul and the metro area network is typically found in the point of presence of the long-haul network within a metro area, where long-haul equipment (e.g., the P-router) links to the metro area equipment (e.g. the PE router). Our consideration of the long-haul network's architectural and technological ingredients is somewhat reductionist. There are at least two reasons for this position.

1. The geographical extent of the long-haul network reduces its link-type diversity to two: (a) long-haul optic fibre plant (cabling and erbium-doped fibre amplifiers) carrying DWDM links and (b) satellite links.
2. The use of energy in long-haul IP routing and optical transport, while difficult to corroborate between the various studies, is likely to be an insignificant percentage of the IPTV VDN's energy use. The GreenTouch project's estimate of the long-haul network's (core IP routers and optical transport) energy use is less than 0.5% of the energy used by worldwide telecommunications networks [66].

The long-haul network serves to interconnect the hierarchy of headends. The IPTV template is illustrated in Fig. 12.

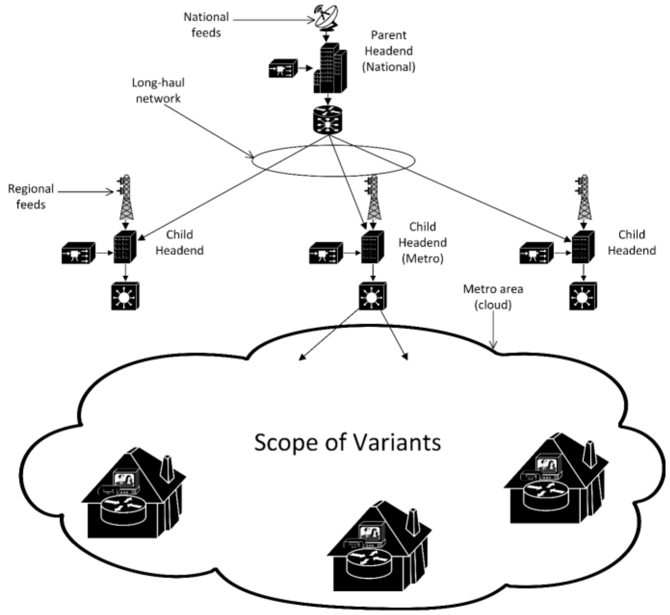


Fig. 12: IPTV template. The scope of the variants is collocated with the scope of the metro area.

The scope of the variants is that of the metro area network. The segments of the end-to-end path that lie within this geographical reach are the primary contributors to the diversity referred to in the introduction to this sub-section. It is therefore apparent that ***the scope for diversity in the IPTV template is the same as the scope for diversity in broadband networks***. We refer to Fig. 13 [170] in our consideration of the variants. This diagram is a ***representative*** illustration of telecommunication networks within the metropolitan area. It does not generalize all possible variations but is representative of the segmentation and the hierarchy in the metro. For example, the Internet Exchange shown is typical of the North American market. In this market, ISPs in a specific metro exchange traffic amongst themselves and with the Tier 1 operator in a specific co-location facility (of which there may be several per metro area). Conversely, the European market for Internet Exchange is better represented by the Amsterdam Internet Exchange (AMS-IX). AMS-IX is distributed among multiple co-location facilities in Amsterdam. At each such facility, ISPs may interconnect with local peers and with Tier 1 providers of long-haul connectivity.

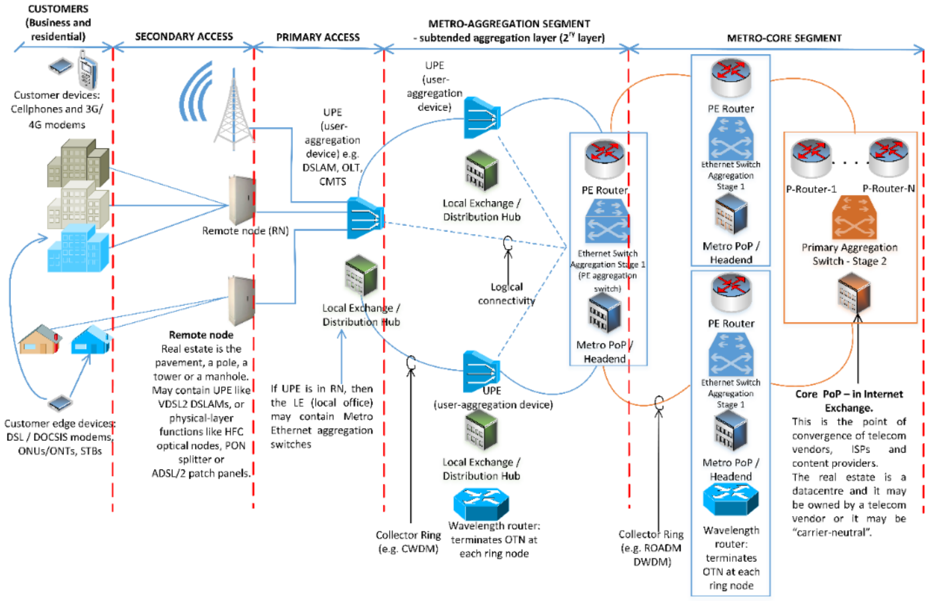


Fig. 13: Representative illustration of telecommunications networks in the metro area [170]

Another significant variation regards the interconnectivity in the metro-aggregation segment. The physical topology of the optic fibre is not restricted to rings (with protection); for example, a single NP's LEs/COs/Distribution hubs may be interconnected in mesh form or in hub-and-spoke form with the metro PoP on the metro-core ring.

Finally, we observe that the IPTV template may be extended to include the metro-core ring. Each PoP on this ring provides up- and downstream connectivity to at least one specific metro aggregation segment. This ring covers a larger geographical extent than the metro-aggregation segment and its topology (with protection) is a good trade-off between resilience and capital expenditure. We may therefore reasonably conclude that the scope for variants lies within the access and backhaul (metro-aggregation) segments and most notably within the access segment. Space constraints do not permit us to explore the various broadband access technologies. We limit ourselves to reflecting upon the way in which video services are driving the evolution of the cable multiple-system operator (MSO) and the diversity that is emerging to provide MSOs with choice in the implementation of their IPTV VDN.

5. Destination: CORD, HERD or Cloudsource?

The iconography in the illustration used in the treatment of the unified headend is familiar to IT operations staff. The treatment itself referred to general-purpose hardware and to software components used to implement video processing functions. This is not surprising; convergence of communications and information technology is familiar. In network providers' social circles, this convergence is shaping the design of the central office (CO) / local exchange (LE; CORD: central office re-architected as a datacentre) and the headend (HE; HERD: head office re-architected as a datacentre).

Netflix's workflow is entirely delegated to AWS except for the CDN. BT has cloudsourced video processing operations from AWS Elemental [152]; BT does, of course, own transport and acquisition infrastructure. These two cases amply exemplify the scalability of cloudsourced operations. Crucially, it illustrates a symbiotic relationship in which huge processing resources are linked to wide-area telecommunication infrastructure and successfully deliver a time-sensitive user application. Since cloudsourcing a service is the alternative to running it on premises, and since COs/LEs/HEs are optimally located (physically) for transmission and switching functions, the justification for CORD and HERD must include prioritisation of these two critical functions. These are the functions which distinguish NPs and which, together with ownership of pathways, cable and right of way, support all NPs' services.

A key tenet in CORD and HERD is software-defined functionality. The unifying concept behind SD-x (software-defined anything) regards the flexibility with which the functional as well as the non-functional requirements can be changed and the efficiency with which these requirements can be implemented. Both flexibility and efficiency need qualification by the specific system's context. With specific regard to use-phase energy, it is good to dispose of methods that facilitate the efficient use of energy in a system. Furthermore, it would be better if these methods, alone, or in combination with others, facilitate changes in the implementation of the system (hence: flexibility) to match a dynamic external environment (i.e. external to the system). The following are some anecdotes that use flexibility in the sense that we are using it.

1. An older reference to the importance of flexibility is found in Sklar's classic [154]. In introducing the advantages of digital communications over analogue communications, Sklar points out that "digital hardware lends itself to more flexible implementation than analog hardware".
2. A reference to the difference between behaviour that is embedded in hardware and behaviour that is embedded in software is made by Jim Larus [177] "[h]ardware is hard, inflexible, produced by gnomes with sub-micron tools. Virtual machines wrap a layer of software around this hardware, and suddenly computers become flexible, malleable and start doing new tricks".

Both these references discriminate between systems along the dimension of flexibility. When considering the development of virtualization on Intel architecture, we can see that hardware has been developed to assist the flexibility afforded by virtualization. This development in hardware has been exposed through Intel VT architectures, which provide the means for the software (hypervisor) to switch between one OS and another. The behaviour of the system changes through collaboration of hardware and software. Development of Intel VT was specifically in response to the need for more efficient virtualization, since VMware's products were using binary translation of Intel binaries to patch critical instructions. Processor utilization efficiency was low. During the life cycle of a VM, the total complement of hypervisor-instructions to support guest isolation is a high percentage of the total number of the sum of hypervisor-instructions, guest-OS-instructions and user-application-instructions. If time spent in memory management for VM isolation is brought into the evaluation of efficiency, then efficiency (without VT-x2) percentages drop further.

Further understanding of flexibility may be obtained through consideration of a hardware-bound function. There are at least two reasons why a function may be hardware-bound. One is that the function is implemented in an electronic or an optical circuit. Another is that the function is implemented in software which cannot be decoupled from its hardware platform. Suppose that this function is required as part of a system's behaviour. Once the function is integrated into the system through interfacing to the hardware item, the location of utilization of the function is limited by the programmability of the paths to and from the interface since the function itself cannot move. Conversely, with function virtualization, the function is implemented on purpose-independent (i.e. general-purpose), vendor-independent hardware.

Rationalization of these anecdotes and examples leads to two conclusions:

1. Flexibility is conferred through ease of re-purposing of an ensemble of functions. Re-purposing can be effected in at least two ways:
   a. moving one or more functions, achieved through function virtualization;
   b. changing the paths to and from one or more functions, achieved through software-defined networking.
2. Efficiency is conferred through hardware support of the new purpose. If this is not to defeat flexibility, then hardware support must be exposed by extending general-purpose architectures.

Function virtualization, software-defined networking and general-purpose processor architectural extensions are the foundations upon which CORD and HERD depend if these are to be the models of future COs and HEs. It seems reasonable to anticipate that the logical culmination of CORD and HERD is a single datacentre architecture in which all functions are implemented on virtual deployment units in a pool of general-purpose servers, under the orchestration of QoE- and energy-oriented video service applications, running on a network operating system.

6. Conclusion

We have addressed VDNs from the perspective of two relevant evolutionary trends that generally characterize the current scenario of communication networks: the growing introduction of virtualization techniques that leads toward increasing network "softwarization", and the constant attention to energy consumption issues. In order to better focus VDN aspects in this context, we have first introduced a classification of them into different genres, based on identifiably different combinations of architectural and technological ingredients. Within this framework, we have briefly examined the main architectural components that play a relevant role, including the headend and the distribution networks, and highlighting their functionalities.

Appendix 4. Data collected through market research and network operator group mailing lists

The raw data about current- and next-generation deployments, that are used to justify claims made in Chapter 8, may be found online[72].

---

[72] https://drive.google.com/file/d/1hqoyTAUOfYpd0FgOgCKpqvbvhvHWx93I/view?usp=sharing

Appendix 5. Questionnaire used in the quantitative analysis of CSP’s implementations of metro area networks.

The questionnaire used to collect the data may be found online[73].

---

[73] https://forms.gle/QtoTkhzEk4Q1BLdVA

Appendix 6 – Screening questions selecting interviewees for qualitative survey

In the following questions, the term "network" is to be interpreted as ***metropolitan area*** instances of that which is described by the Broadband Forum's document TR-145, between T and A10, as illustrated in Fig. 1, i.e., between residential/commercial end users (subscribers) and services within the metro area.

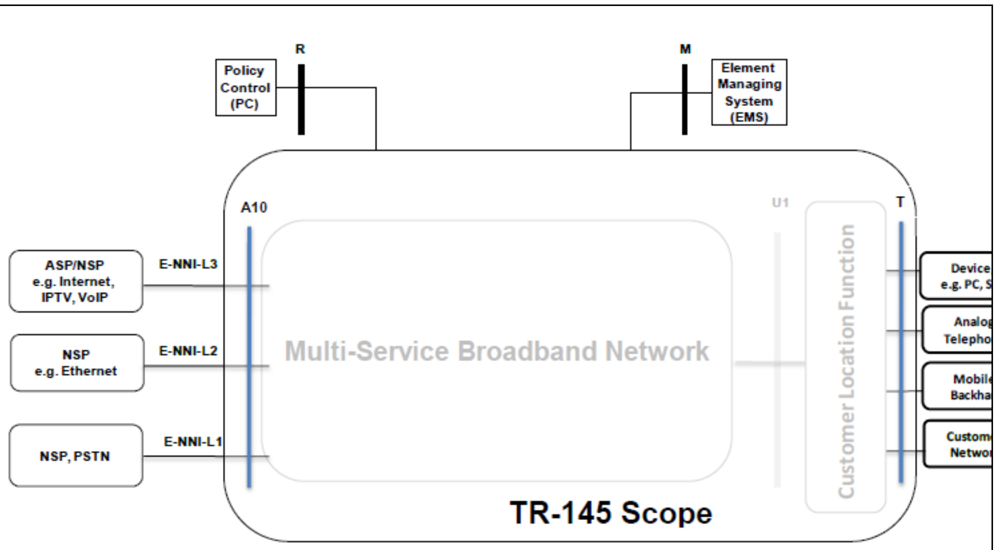


Fig. 1: Scope of the term network in this set of questions = TR-145 Scope

1. What country/region have you ***deployed/operated*** networks in? [Select one]
    - North America
    - Western Europe
    - Central Eastern Europe [Screen Out]
    - China [Screen Out]
    - Middle East [Screen Out]
    - Emerging Asia-Pacific [Screen Out]
    - Developed Asia-Pacific
    - Latin America [Screen Out]
    -

2. What is your job title? [open-ended]
    - Network planning: Network engineers
    - Network engineers might also be known as:
        - (1) "network design engineer"
        - (2) "network planning engineer"
    - Network operations: NOC (network operations centre) engineers
    - NOC engineers might also be known as:
        - (1) "operations engineer"
        - (2) "technical operations manager"
        - (3) "headend engineer"
    - Network operations: Field engineers
    - Field engineers might also be known as:

    - (1) "network technician"
    - (2) "network maintenance technician"
    - (3) "maintenance technician"
    - (4) "field technician"
    - (5) "supervisor, field operations"
    - (6) "supervisor, technical operations"
    - (7) "communications technician"
- Other related role: ______________
    - Please specify how it is related ____ ____ ____ ___

3. Please, indicate, as you deem fit, one – two items of operational and/or engineering experience/projects/accomplishments [open-ended]
    - ____ ____ ____ ____ ____ ____ ____

4. Please specify the type of network operator **you** work for (select all relevant categories).

Check all that apply.

☐ Fixed-line telecommunications operator (classical telco)
☐ Mobile network operator (MNO) [Terminate]
☐ Cable operator (or multiple system operator - MSO)
☐ Wireless Internet Service Provider (WISP) [Terminate]
☐ Virtual network operator [Terminate]
☐ Other: ____

[Terminate if only MNO or only WISP or only MNO & WISP]

5. Do you have any responsibility for the following areas? [Select all that apply]
   - ○ network planning
   - ○ network operations
   - ○ I have no involvement in network planning or network operations at my organisation [Screen Out]
6. Which of the following best describes your responsibility and involvement in the (metro area of telecommunications) networks that you work with? [CHECK ONE]
   - ○ I make the decision, solely or jointly, regarding the metro areas network in my organisation
   - ○ I have significant influence over the metro areas network in my organisation
   - ○ I have some influence over the metro areas network in my organisation
   - ○ I have no influence at all
   - ○ I do not know

Appendix 7: Power use in cloud-native video streaming

1. Introduction

Content delivery networks (CDNs) are overlay networks that are key to controlling the growth in demand for bandwidth in long-haul communications links. By distributing content to caches in geographical regions of the world where customers are located, the number of times which a single item of content crosses long-haul links between the content origin's region and the customer's region, is reduced to just one. In turn, the content is distributed several times to customers in the region. While the function of the CDN, from the customer's perspective, is that of reducing latency and avoiding buffer underrun, the control of bandwidth growth is a function that has a strategic role in the stability of world-wide communication. The CDN's role in bandwidth control continues to gain attention [360]; a variety of CDN implementations has been investigated [361], [362] and surveyed [363], [364] and generalized surveys are of ongoing interest [365], [366]. The importance of the CDN seems to grant sufficient ground for study of the impact of its point of presence (PoP) on the information and communication technology of its environs.

The architecture of the CDN PoP comprises a variety of sub-systems. The foundation consists of **storage systems** that serve as the geographically localized cache of the content's origin. These storage systems are an offline store that retains the content in the event of power loss and is refreshed, by a **content delivery application**, according to the CDN's policy for updating cache. For example, the content delivery application can use known lulls in the locality's activity to replace expired content with its update. If the local storage systems' cache is incomplete and suffers a miss, the application can pull content from origin nodes. Two well-known policy variants are implemented in push CDNs (the origin proactively pushes content) and pull CDNs (content is demanded on first miss).

In addition to these storage systems, local caching to temporary storage (such as volatile random-access memory (RAM)) might be used to reduce latency of access. Algorithms run on these **caching servers** to determine which content to evict from RAM when insufficient space exists for requested content deemed a candidate for RAM. Caching servers may also serve as **media servers**; the media server transforms the video file into a stream which is delivered to the video player operated by the video consumer. The process of transformation may include encoding or transcoding, and it may include adapting to current bandwidth availability between media server and video player (adaptive streaming). Media servers are protected from malicious overload by Distributed Denial of Service (DDoS) perimeter systems. These systems detect characteristic behaviour of DDoS attacks and intervene to block their development to full scale.

While IP anycast systems support distribution of demand between geographically-distributed CDN PoPs, the **load balancer** subsystem of the CDN PoP supports the distribution of demand at the local level. Local demand is balanced between existing instances of media servers. Balancing proceeds among existing instances, until utilization thresholds are crossed. If a lower threshold is crossed (downward), one or more instances can be decommissioned. If an upper threshold is crossed (upward), one or more instances can be brought into service. This agility is supported by the use of virtualization containers (VCs - whether VMs or containers, also referred to as virtual entities (VEs)). In particular, containers' low startup time is well suited to this "smart standby" [47], [367] mode of operation; indeed containers' startup time is very much shorter than virtual machines' [368]. Thus, smart standby of containerized media servers is a valid component of the overall approach towards minimization of the CDN PoP's power use.

This study seeks to compare power use in containerized deployment of the media server in a CDN PoP. The study comprises two parts. The first part focuses on the power use of the media server as it processes a representative set of tasks. The media selected for study is video (henceforth, the media server will be referred to as the video server), and two reasons support this choice. Video dominates traffic, whether in the access, aggregation, metro-core, or long-haul. Moreover, some of the tasks, such as transcoding, are processor-intensive and serve to indicate the power capacity required to support CDN PoPs. The second part complements the first with an investigation of a virtualized network function (VNF), namely, layer 2 switching. This supports understanding of scenarios in fixed fifth generation (F5G) networks, as the latter participate in the paradigm shift towards virtualized entities (VEs/VCs).

2. Background

2.1 Power models

A better grasp of the impact of containerization on a CDN PoP's power use, requires understanding of containerization's impact on the media server's use of power. Media servers are deployed on general purpose computing systems (or: commercial-off-the-shelf computer systems (COTS)), and a basic understanding of these systems' power characteristics must support further study.

Power used by a computer system has an idle (static, or leakage) part, an active (dynamic) part and overhead. The power use referred to here is a **system** metric: it is an aggregate that sums all consumers' (system components) power use, whether it be of dynamic, static or overhead type. Idle power's (P_{idle}) usefulness depends on perspective. On one hand, idle power use is a real and significant cost: from the perspective of facility managers and sustainability advocates, it is of interest. On the other hand, it is irrelevant to the way in which processes exercise a computer system: it is of secondary importance in a study of the power required to deliver a service. It follows, then, that a study of the power used to operate a CDN PoP's server systems must primarily engage with **service** power use, or, using general terms, with dynamic power use.

A simple, yet useful, classification of power models divides them into two: (a) one that treats the computer system as a black box, and uses workload to predict power in real time, and (b) another that exploits knowledge of microarchitecture and architecture [369]. They represent two different levels of abstraction (see [222, p. 42] for a more detailed distinction) of a computer system.

2.1.1 The affine relationship between aggregate power and utilization

The affine relationship is a well-known exemplar of the black-box class. A typical representation of workload may be one or more parameters of utilization (e.g., MIPS and IOPS (input/output operations per second). The affine relationship between power use and utilization [369] is well suited to describing legacy network equipment [370]. The general form is reproduced as equation (1).

$$P(\rho) = P_{idle} + \big(P(\rho = 1) - P(\rho = 0)\big)\,\rho \quad \text{.......} (1)$$

where $P(\rho)$ expresses power at utilization ρ.

Equation (1) expresses power in terms of a generalized utilization, but particular forms like processor load in millions of instructions per second (MIPS), or switching throughput in bits per second may be used (where the model holds true). Note that equation (1) refers to the idle power ($\mathrm{P_{idle}}$), but not to the overhead. The static part and overhead are significant and cannot be ignored. However, idle power use has no correlation with the computer system's load. Furthermore, while the overhead (such as fan power use) can indeed be expected to relate to load (it is not a constant type of overhead), power used by these overhead drains can be expected to have much longer response times than that of power used by silicon. For example, a fan's speed will increase when the temperature in a thermally instrumented zone increases; heat capacity is clearly a factor that will affect temperature rise, as well as temperature drop. Therefore, fan speed will not follow silicon loading and inclusion of fans' power use in measurements will obfuscate the dynamics of power use by silicon under load.

2.1.2 Limitations of the affine relationship

Accuracy of the affine relationship has been shown to worsen when the processor does not dominate dynamic power [369]. Apart from processor utilization, system power models have been developed to handle other system components like primary (silicon dynamic random-access memory) and secondary storage (disks) [371]. Furthermore, processor power and frequency are quadratically related [372]. One approach to handling frequency variability is given in [373], where the affine relationship is modified and takes the form shown in equation (2):

$$P(\rho) = P_{idle}^{(f)} + \left(P^{(f)}(\rho = 1) \ - \ P^{(f)}(\rho = 0)\right)\,\rho \quad \text{.......} (2)$$

In equation (2), the frequency index in $P_{idle}^{(f)}$ and $\left(P^{(f)}(\rho = 1) - P^{(f)}(\rho = 0)\right)$ serves to denote the dependence of intercepts (static/leakage/idle power) and gradients on frequency of operation. Evidently, the affine model expressed in equation (1) does not describe a computer system's processor's power use when the processor is operating under dynamic adaptation of voltage and frequency (dynamic voltage and frequency scaling (DVFS)). However: system power measurements must be adjusted by a baseline that includes $P_{idle}^{(f)}$. A corresponding measure will be dealt with in considerations of measurement methodology.

2.1.3 Relevance of the architectural models

Power architectural models predict power use as a linear function of several activity indicators. These indicators regard the activity of aspects of architecture and microarchitecture of a system. Therefore, models that harness microarchitectural activity indicators tend to be bound to specific hardware models. Activity indicators are commonly referred to as performance counters. Architectural models are well suited to the task of measuring dynamic power use. Performance counters compiled by the operating system are architectural; those compiled by the hardware in dedicated registers, are microarchitectural. System software can abstract microarchitectural counters by a layer that returns these counters through method calls.

A particularly useful class of these counters obtains power use directly. Examples include Intel's Running Average Power Limit (RAPL) and AMD's AMD Energy Driver (amd_energy). The feature addresses power use through hardware support and can form part of a measurement methodology. For example, RAPL's MSR_PKG_ENERGY_STATUS register provides a running, cyclic total of energy used by the CPU package (all cores included).

2.2 Isolation and attribution of dynamic power use

Dynamic power is used during both service idle time and service delivery time. The power used during service idle time is **not** the P_{idle} of equation 1. Rather, it is the dynamic power used by system software (whether in user or kernel mode) in order to maintain system operation. Service idle time's power use will be subtracted from service delivery (during video streaming) time's power use, in order to obtain the differential relevant to this research. Service idle time's power use is a tangible justification of the requirement to use minimalist general-purpose systems. Since user application and system software processes and threads are many, then minimization of such root causes simplifies the process of attribution of dynamic power use. An illustration of the multiplicity of subsystems of the computer that are in the scope of power use measurement may be found in [374]; clearly, detailed inspection is required to correctly isolate and attribute power use. Two broad classes of process and thread are identifiable, and are briefly described below. Sub-sub-sections 2.2.1 and 2.2.2 concern the video server, but the same considerations readily hold for the virtual switch's server too.

2.2.1 Kernel operations

There are several categories of operation carried out by the kernel to support the operations of the video service. These include:

- processing of hardware interrupts when packets arrive, and concomitant activities like the onerous requirement for the system call to return to userspace;
- managing the timer, to schedule processor allocation to processes and threads;
- memory and cache management, and
- processing of system power monitoring instructions.

The detail of which processes to monitor is expected to be captured in the baseline (see approach-baselining, below). The power used during service idle time will be subtracted from power used during service delivery (during video streaming), in order to obtain the differential relevant to this research.

2.2.2 Service operations

The video streaming service may be tersely described as one in which:

- a source file is encoded (or transcoded) using a video codec and an audio codec;
- the codecs' output is packetized and
- transmitted over a network interface as the payload of a communication protocol that handles:
  - the correct sequencing of the received content (payload) and
  - adaptation of the video quality of the content to network conditions.

These operations must be matched to specific computing entities (components such as processes, threads, main memory and cache) in the computer system and the power use thereof is to be monitored. In particular, the specific computing entities are expected to include the video server process obtained by running the principal executable, and library functions which it calls to support the three major categories of operation listed above (encoding, packetization and sequencing into a stream of adaptable quality).

2.3 Service scaling

Service scalability is essential to cope cost- and energy-efficiently with short-term fluctuations in demand. These fluctuations are commonly referred to as the daily diurnal and nocturnal crests and troughs In Internet service demand. **Fig. 1** [375] illustrates service scaling in a virtualization infrastructure. The range of service supply varies from minimum scaleLevel to maximum scaleLevel, stepping with the size corresponding to a virtual network function component (VNFC). Higher demand (load) can be met by spawning one service instance per client. The service instance may consist solely of a single VNFC.

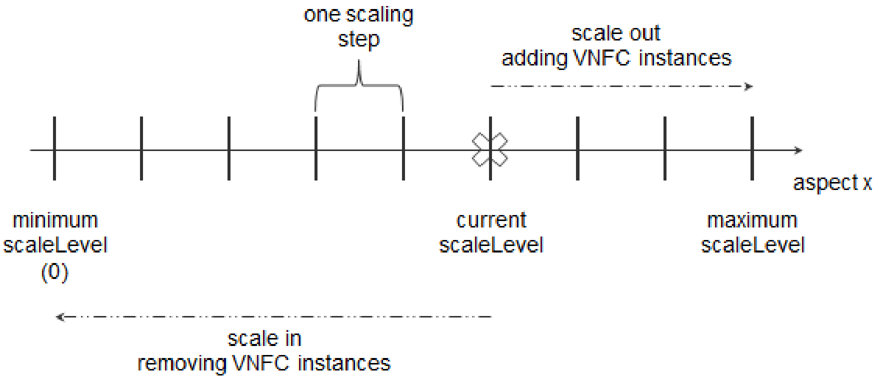


Fig. 1: *[375, Fig. 5.1–1]*: **Demand is met by deploying over a range from minimum to maximum scaleLevel**

3. Objective

3.1 Principle

An overhead is expected in the containerized implementation, and its ***quantification*** is sought. The objective can be articulated in terms of a comparison between two types of deployment:

- power use in a computer system that runs the service within containers, with
- power use in a computer system that runs the service directly on the operating system.

Quantification is sought in order to control a tradeoff between native and containerized deployment. The tradeoff may be succinctly summarized as one of greater operating power per unit (physical host) versus potential for lower number of operating units (physical hosts). The following sub-sections elaborate on this summary.

3.2 Greater operating power per unit

3.2.1 Quiescent operating power

A host (physical server) computer system uses power in its quiescent state. Quiescence is the condition where the host has an active operating system and is running a minimal set of services. Levels of quiescence can be defined, in accordance with different specifications of the set of services. In all levels, quiescent power use consists of an idle/leakage/static component, due solely to physical properties of the hardware, and a dynamic component, due to execution of software processes on the hardware. A containerized deployment uses more power in the quiescent state because its minimal set of services is a superset of that used by a native deployment. Therefore, even while no video clients are served, a containerized deployment has greater operating power. Moreover, in order to grasp the difference between operating power of the two deployments, a service process deployment strategy has to be defined.

3.2.2 Full load operating power

As clients appear, service processes must be started to handle the workflow. Minimally, the video service workflow consists of the following cyclical process:

(a) fill a memory buffer queue by copying some initial large chunk of the file from storage;
(b) transmit the queue head;
(c) repeat queue head transmission until a fraction of the queue is empty;
(d) re-fill the memory buffer queue from memory ramdisk and
(e) repeat steps (b) to (d) until all the file has been read into the queue.

The process, which can be tersely summarized as streaming, is independent of the file's encoding format, but will be extended should real-time transcoding[74] be necessary to meet the client's constraints. These observations prompt the identification of **load units**, comprising the full amount of work (in Joules) required to process the workflow. A topmost classification divides the set of load units into two branches: one for the case where only streaming is needed, and another for the case where both streaming and real-time transcoding is needed. Below this topmost classification, load units can be identified for every encoding type and bit-rate preset. Each such load unit corresponds to a single resource unit, which is the bundle of computing and networking resources required to serve the load unit. Specification of a load unit supports the analysis of full load operating power, as the latter is used when no more load units can be taken (subject to some quality of service (QoS) condition, as described below). This limit can be articulated better in terms of bin packing, where each physical host is represented as a bin capable of serving load. When a load unit is served, the bin is partially filled and the corresponding resource unit is removed from the aggregate of the host's available resources. As more load units are served, the bin is progressively filled until no more load units can be added. This is full load, and power used under this condition is the full load operating power.

3.2.3 Bin packing

The condition of full load corresponds to the operating principle of maximization of capacity utilization without degrading key indicators of quality of service (QoS). That is, if each server represents a bin of some service capacity C, then the server is loaded until its capacity is fully utilized without degrading the QoS. The process of filling the server suggests modelling using bin-packing algorithms; hence, depiction of the server as a bin.

Since both capacity, C, and QoS are complex, a simplification is sought to manage the tractability of the problem. Let the capacity, C, be the number L of load units U that a host H in a set of homogenous hosts, can serve without degrading the received bit rate at any client, below the preset for

---

[74] Anticipatory and real-time transcoding may be distinguished. The former prepares various video files, one per bit-rate, and possibly different compression formats. The latter changes rates and/or compression format when it is unavailable in storage.

U. This specification of C and QoS key performance indicator (KPI) serves to support specification of other, different loading conditions for both types of deployment. This is reserved for future study.

3.3 Potential for lower number of operating units

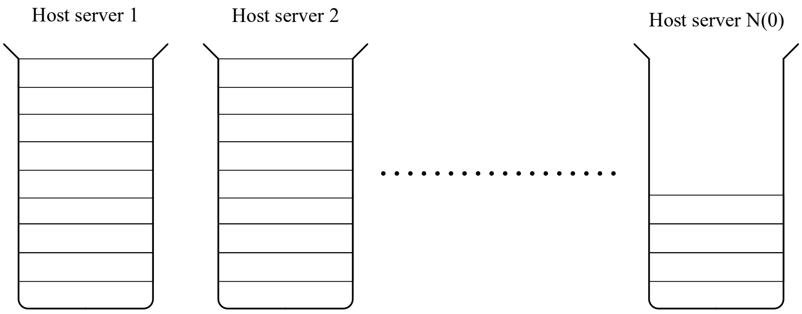
Consider the condition of consolidation, obtained by deploying video service process to the minimum number of hosts possible. **Ideal** consolidation is obtained when all N hosts (bins) in service except for the Nth are packed. Here, bin-packing corresponds to loading a server until the bit rate served at one or more of the clients falls below the preset.

Such an idealized consolidation is depicted in Fig. 2. The top part (a) shows ideal consolidation, at some time t = 0. $N(@t = 0)$ (henceforth denoted by $N(0)$) servers are shown, of which $N(0) - 1$ are filled and 1 is partially filled. One white segment represents one utilized resource unit. One context in which this consolidation is achievable is when an initial set of load units is presented to a dispatching subsystem for distribution onto a set of idle servers. This context applies to both the case where the service is running as a user application (UA) directly on the host OS (henceforth shortened to "running as a UA") and the case where the service is running in a container.

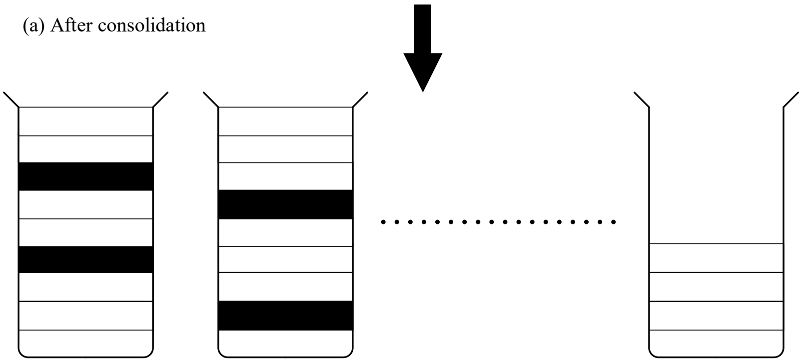
Over time, clients drop out (the black gaps represent unutilized resource units) as their viewing sessions terminate. While running as a UA, the service instance supporting dropped clients terminates and leaves a resource gap. However, these gaps cannot be filled with running instances on other servers, since UA state cannot be migrated as easily as when it runs within a container. At some arbitrary time, t, after service starts, it may not be possible to consolidate the service running as a UA, but it should always be possible to consolidate the service running containerized. Therefore:

- while server $k \in \{1, 2, \dots N(0)\}$, draws $P_k^{(c)} > P_k^{(ua)}$, where $P_k^{(c)}$represents power drawn while serving a capacity-sized subset from containers and $P_k^{(ua)}$ is the native counterpart, and
- while $N^{(c)}(0) \geq N^{(ua)}(0)$, since $L^{(c)} \leq L^{(ua)}$, where $L^{(c)}, L^{(ua)}$ are the respective capacities of the containerized and native service deployments,

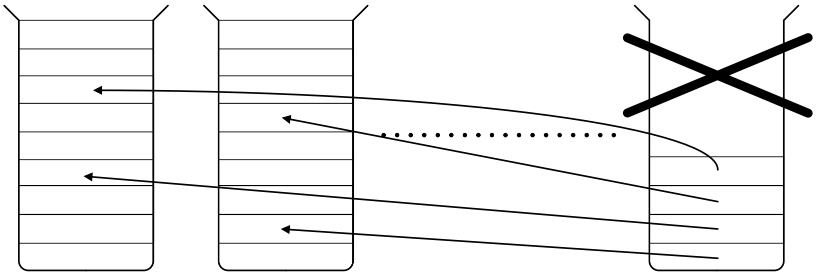
there is no predetermined relationship between $N^{(c)}(t)$ and $N^{(ua)}(t)$. Moreover, while all but the last of the $N^{(c)}(t)$ hosts can (at least periodically) be subjected to consolidation and thus use power amounting to $P_k^{(c)}$ W, the operating state of the $N^{(ua)}(t)$ hosts, and their individual power uses, is unknown. It thus follows that the relationship between $P_{N^{(c)}(t)}^{(c)} + \left(N^{(c)}(t) - 1\right)P_k^{(c)}$ and $\sum_{k=0}^{N^{(ua)}(t)-1} P_k^{(ua)}$ is not evident, and **quantification** of the overhead $P_k^{(c)} - P_k^{(ua)}$due to containerization, is a necessary prerequisite to understanding the scale and usage pattern at which containerized deployment is energy efficient.



(a) After consolidation



(b) After clients leave (native deployment)



(c) After clients leave and containers are moved (containerized deployment)

Fig. 2: Simplified view of power control enabled by containerization of service application

4. Implementation model

4.1 Overview

An edge cache of a video streaming service is deployed. A high-level view of the implementational model is shown in **Fig. 3** and **Fig. 4**.

- **Fig. 3** shows an implementation that is easily portable to a cloud-native infrastructure (henceforth referred to as the cloud-native implementation), while
- **Fig. 4** shows an implementation that is a hybrid of physical (the video server) and virtual network functions (the switch).

The cloud-native implementation (**Fig. 3**) uses containers to host both the video server and the video player. A virtual layer 2 switch is hosted in the intermediate node. The intermediate node hosting this VNF is described in [337] as a subtended access node; this latter type of access node supports better performance for a service area, and reduces primary feeder cable lengths to the service area, at the cost of greater complexity (several subtended access nodes are required to cover the upstream access node's subscriber base). The VNF switches packets between the media server's port (mostly incoming) and the port to which client devices connect. The hybrid implementation runs instances of the video server as UAs on the host operating system, and the same layer 2 VNF as the cloud-native implementation. The use of a container for the client is only for convenience's sake, to facilitate scaling of the client and generation of higher load. Spawning of new client instances is facilitated by the testbed's infrastructure at the subscriber's end. From the perspective of the video service provider, there is no direct impact on service delivery, as long as the streams are unicast to the respective subscribers.

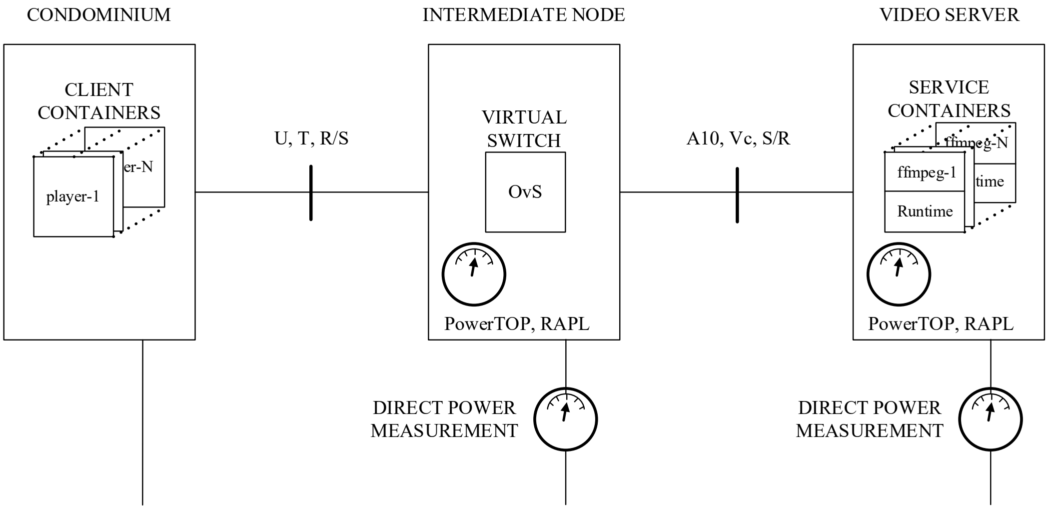


Fig. 3: Physical topology of the video streaming service, deployed in containers. Video Server located in local exchange or Access Node and Intermediate Note located in street cabinet.

4.2 Metro-area network topology, technology and reference points

The video player's host connects to the Communications Service Provider's network through an "Active Ethernet" [376] access network. The client end is labelled "condominium" to illustrate the deployment when scaled to public service. In dense urban areas, the IN may serve several large multi-dwelling units (MDUs, or condominia); each MDU may be served by a single port on the layer 2 VNF device, with local switching within the MDU (not shown) distributing towards individual subscribers.

Reference points are shown in in **Fig. 3** and **Fig. 4**. In Active Ethernet, the U and T reference points (RPs) coincide at the video player's (subscriber's) end (there are no technology stack adaptation functions needed to match the CSP's access network to the subscriber's network). When optic fibre is used, the R/S RP too coincides with the U and T RPs. Moreover, the A10 and Vc RPs coincide downstream of the access node (where the edge cache is located). When optic fibre is used, the S/R RP coincides with the A10 and Vc RPs.

This implementation is a realistic testbed for a video streaming service that locates its edge cache in a local exchange (the access node), with point-to-point connectivity between the exchange and the subscribers. Such a scenario has been shown to be the largest ***by number of operators*** [376] among respondents from network operator groups around the world.

The second implementation (**Fig. 4**) retains the virtual switching device but deploys the video server and the player directly on the host operating systems. The second implementation will provide the reference against which to compare power use of the video server while running as a containerized application, with power use of the video server while running as a UA on the host operating system.

4.3 Hardware

The hardware used in this testbed consists of a set of HPE Gen9 BL460c blade servers, hosted in an HPE c7000 blade enclosure. Connectivity between the blades is obtained through pass-through interconnect bay modules, patched with single-mode optic fibre cables. These latter modules enable a bypass of physical network switches. Bypass is necessary to introduce separate switching hardware, in support of the goal of use of networking devices outside the c7000 ecosystem's range. The virtual switch is implemented on a third HPE Gen9 blade server.

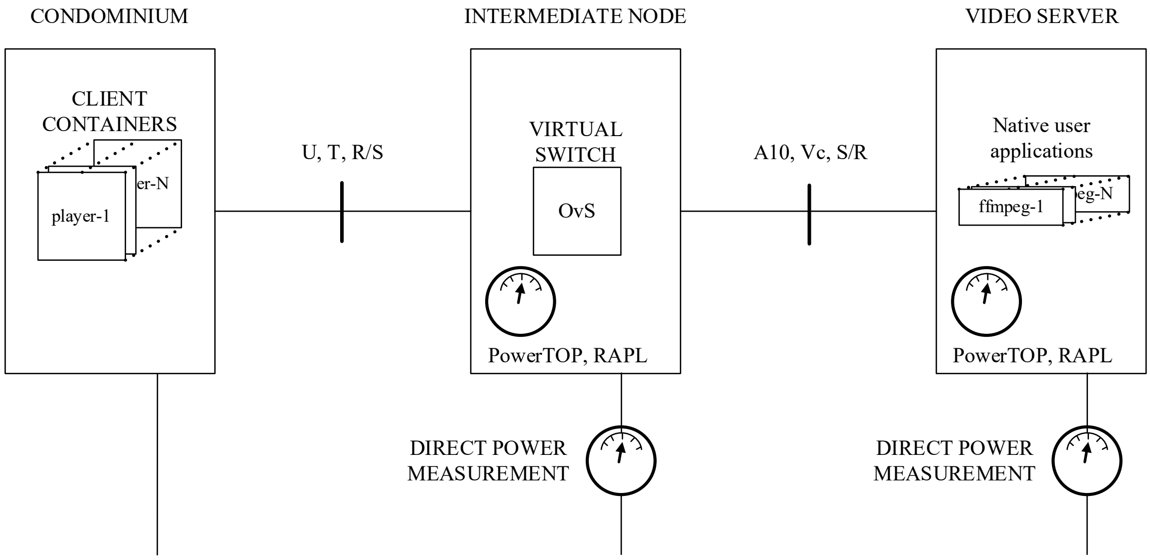


Fig. 4: Physical topology of the video streaming service, deployed on a host operating system

4.4 Software

The software consists of:

(a) an FFmpeg video server. This runs on a Gen9 blade, and is representative of the access node at the edge of the metro-core network;
(b) an FFplay video player. This runs on a Gen9 blade, and is representative of the player used by end user resident in an MDU, and
(c) the virtual switch software is Open vSwitch. This runs on the third Gen9 blade in the test hardware.

The operating system was selected on the basis of simplifying isolation and attribution in power measurements. Isolation and attribution (see sub-section 2.3) are simplified if the operating system is reduced to the minimum necessary to operate the (video streaming) service. This condition does not impinge upon the realism of the testbed, as minimization of software footprint is good practice, at least for security purposes (minimization of attack surface). Moreover, it is evident that minimization reduces the demand for secondary storage capacity, as well as for primary storage. While minimalist operating systems do not necessarily correlate with minimal noise in power measurement, it seems useful to reduce the number of possible sources from the outset. For this reason, Alpine Linux Standard distribution was chosen to support UAs and containers. For example, Alpine Linux uses musl libc instead of GNU libc to minimize its footprint, and uses BusyBox in further support of the goal of suitability to resource-constrained environments (e.g., embedded systems). While the environment in this case has plenty of resources, it is expected that the minimalist operating system will be found to have a lighter impact on power use.

The container system software selected is Docker. Docker is a mature containerization platform and it is modular: the runtime daemon (containerd) supports other user interfaces apart from the Docker user interface (dockerd). For example, Kubernetes can be used to manage containers created through the Docker CLI.

5. Measurement methodology

5.1 Instrumentation

5.1.1 Aggregate power use

Selectivity in aggregate power use measurement is afforded by blade systems, since these separate power supply to the (blade) computer system from power supply to two major overhead power drains. Blade servers use blade chassis services for power supply (where ac – dc conversion losses occur) and cooling (where blowers use power as they ventilate from chassis front to chassis rear). Thus, measurement of power used by the blade server at the supply voltage rails is free of the problematic, variable contribution from overheads, and idle power can be measured to the accuracy afforded by these blade system power measurement instruments. Summarizing:

- blade instrumentation avoids the overhead corresponding to losses in the power supply and blowers, and
- the measurement provides an envelope which serves as a strict upper bound against which to validate readings obtained through more granular methods.

The blade system's measurement, henceforth referred to as the iLO[75] measurement, is of integer type.

---

[75] iLO stands for "integrated lights-out". The iLO sub-system is a management module that provides the system administration with remote management and instrumentation of the server where it is integrated.

5.1.2 Granular power use

The desired granularity of measurement is the process and thread level. This granularity reveals the dynamic power use made by containers, and therefore suffices for this work's purpose. The running average power limit (RAPL) feature addresses power use through hardware support supported in the blade's processors. However, RAPL does not inherently attribute power usage to processes. Attribution requires that the activity of processes and threads be monitored. Therefore, granular measurement requires software as well as hardware support.

Software can obtain power use either from the bottom upward (by counting power difference between the start and end of a process or thread's time slice on a core) or from the top downward (by dividing the power measurement over a period amongst processes and threads in proportion to their core utilization). Since this specific feature – i.e., attribution to processes and threads – is indeed supported by PowerTOP [377], then this tool complements the aggregate power measurement obtained by blade sensor instrumentation. PowerTOP uses a top-down approach [378], and precedes the measurement period by one of calibration in which it obtains weighting parameters for the attribution process. Calibration is further refined with use, and PowerTOP saves its parametric refinement to persistent storage for future exploitation [379].

PowerTOP does not report the power used by hard disk drives (HDDs). Power used by HDDs must be accounted for separately.

5.2 Baselining: separating dynamic power uses

This section sketches the details of how dynamic power use during service active time is distinguished from dynamic power use during service idle time. The process is referred to as baselining.

Notionally, a baseline is a reference observation which pervades all subsequent observations, as a component thereof. Here it is helpful to think of the baseline as an ordered set of reference observations, each observation therein encapsulating preceding observations.

(a) The first baseline is that of power used while the blade runs the minimalist operating system. This corresponds to $P_{idle}^{f_1} + P_{quiescent}^{(OS)}$, where $P_{quiescent}^{(OS)}$ is the dynamic power corresponding to the server's background operation **without** container system software and idle service, and $\mathrm{P}_{\mathrm{idle}}^{\mathrm{f}_1}$ is the static (idle/leakage) power used at the frequency of operation of the quiescent state. Instrumentation:
  - blade power sensors
  - PowerTOP

(b) The second baseline is that of power used while the blade runs with the container runtime daemon and the docker daemon active, but without active or exited containers. This corresponds to $P_{idle}^{f_2} + P_{quiescent}^{(OS+dockerd+containerd)}$, where $P_{quiescent}^{(OS+dockerd+containerd)}$ is the dynamic power corresponding to the server's background operation **with** container system software and idle service, and , and $P_{idle}^{f_2}$ is the static (idle/leakage) power used at the frequency of operation of the quiescent state. Instrumentation:

- blade power sensors
- PowerTOP

Note that both $P_{quiescent}^{(OS)}$ and $P_{quiescent}^{(OS+dockerd+containerd)}$ are dynamic power use.

5.3 Mitigating errors

The principal source of error is measurement uncertainty at the iLO, as the iLO rounds to the nearest integer. Since the iLO rounds $[n - 0.5, n + 0.5)$ to $n, \in \mathbb{N}$, then, without further information on the probability density function (pdf) of the error, a fair representation of each measurement is the value n obtained by the iLO. This contrasts with the floor (round down/truncation) function, where a fair representation of a measurement n would be $n + 0.5$, or the ceil (round up) function, where $n - 0.5$ would be fair. Of the three conversions from real to natural number representation, rounding to the nearest integer has the least maximum error, and this corresponds to 0.5 W.

The ideal statistical distribution of errors is that of a uniform probability density function. If measurement errors were indeed so distributed, then the mean of actual measurements can be obtained as the mean of the set of errored measurements. However, for the specific operating context of a quiescent operating system, the probability of a non-uniform distribution cannot be neglected because the dynamic power is low enough to keep the total power's range within half a watt. This is prone to persistent positive bias in error or persistent negative bias. In such non-uniform pdfs, the actual mean cannot be obtained; only a range of values within which the actual mean lies, can be obtained.

Both baselines regard quiescent states. If bias is detected, mitigation can be pursued through the less biased of the two baselines. The better baseline can be used to compute the affected baseline as the arithmetic combination (addition/subtraction) of the better baseline and the difference in dynamic power between the two baselines. Therefore, each measurement of baseline power must be accompanied by a measure of dynamic power, to support evaluation of the error in the means obtained through the iLO's measurements.

This approach notwithstanding, it may still not be possible to reconcile the two baselines in this manner. In such an eventuality, the ranges of values within which the actual means lie can be combined with the difference in dynamic power between the two baselines. The objective remains that of reconciling all measurements, within the margin of error anticipated.

5.4 Video service operation

5.4.1 Conceptual framework

Video service will be delivered from both containerized and native deployments (see objective). The test conditions pertinent to the video server will be the following.

1. Deployment
    a. During containerized operation, each video service process and the libraries on which it depends will be operated from a container. One service process serves one client.
    b. During native operation, a new instance of the video service process will be started for every new client.
2. Load unit: This will consist of the work required to process a workflow based upon a video with the following technical specifications:
    a. Overall bitrate = 457 kb/s, = video bitrate of 326 kb/s + audio bitrate of 127 kb/s + mp4 container metadata rate (overhead)
    b. Duration = 01:32:02.19 (5522.19 s), of which 30 minutes are played.
    c. H.264 video codec, Main profile
        i. Resolution = 1280 x 720
        ii. Frame rate ≈ 23.98 frames/second (fps)
    d. AAC audio codec, Low Complexity profile
        i. Sampling rate = 44.1 kHz
3. Client supports same video and audio codec; no real-time transcoding necessary.

An **operational profile** will be obtained through a sequence of experiments that exercise the video server at progressively higher levels of workload. The profile will be captured by operating the video server through both the containerized and the

1. **Level 1:** The video will be streamed to the client and power use compared with that used in baselines 1 and 2.
2. **Level 2:** The video will be streamed to several clients, at periodic offsets from one another. Power use will be compared with that used in level 1, and with that used in the baselines.
3. **Level 3:** The video will be streamed to as many clients as the video server is capable of serving while meeting the bitrate demand of each stream, with streams starting at periodic offsets from one another.

Instrumentation involved here will be:

- blade power sensors
- PowerTOP

The difference between the operational profile's power use in the containerized deployment and native deployment, comprises the objective identified earlier.

5.4.2 Experimental procedure

The power used by the video server and virtual switch is measured at various scale levels. An instance consists of a container carrying ffmpeg. A single container is created to deliver a single stream and is destroyed immediately thereafter. When the container is created, ffmpeg is executed and listens on a TCP port, through which it streams the video (duration = 1:32:2.19 h) until termination.

Management of operations is not trivial, even at the minimum scaleLevel, as it involves the following steps:

1. Reboot the video server, to obtain a common and reproducible initial state.

2. Wait until the video server quiesces. This is the time required for server power use to fall to the state where the iLO measurement persistently shows baseline 2 usage.
3. Start the power meters for both total and dynamic power, for both the video server and the virtual switch.
4. Wait for a fifteen-minute interval, to capture behaviour before video streaming.
5. Start a container carrying the ffmpeg listener.
6. Start a client to connect to the container.
7. Wait for a fifteen-minute interval, to capture behaviour after video streaming.

For several concurrent streams, steps 5 and 6 must be repeated for each one of the additional streams. It seems evident that manual management is highly prone to error and is therefore unsuitable. Automated management using Ansible is employed to handle the **orchestration** of the various roles: power meters, container runtime managers and video clients. This enables the experiment to be scaled out to levels that are well beyond the physical limitations of a single human operator.

Measurement of power use of a single instance is first made (this is the minimum scaleLevel – see **Fig. *1***) through the manual approach, to validate the procedure's steps. Power use with containerization is compared with power use by the native service. The procedure is then compiled in the format of Ansible playbooks (see Appendix 2) and used as the basis for subsequent scaling outwards.

Automated operations are started by repeating operations at the minimum scaleLevel. scaleLevel is then progressively increased to obtain power used over a wide load range.

6. Results

Denote:

- mean dynamic power measured by PowerTOP by $\overline{p_{dyn}^{(ptop)}}$
- mean total power measured by the iLO during a time period T_x by $\overline{p^{(iLO)}}[T_x]$.

6.1 Video server's Baseline 1: $P_{b1}^{video}, = P_{idle}^{f_1} + P_{quiescent}^{(OS)}$

6.1.1 Power measured by the iLO's instrumentation

Since the iLO truncates decimals in $[n, n+1)$ to n, then the computation of the mean will count the incidences of 45 W and 44 W, and use them as weights to compute a lower limit to the range of values which the average can take. An upper limit is obtained by adding the maximum possible error (equal to 1) and the mean of the possible range obtained by adding the mean error (0.5) to the lower limit of the range, Using this premise, mean power measured by the iLO, under the condition of a quiescent operating system (see Fig. 5, Table I):

$$P_{idle}^{f_1} + P_{quiescent}^{(OS)} = \overline{p^{(iLO)}}([10{:}31{:}49, 11{:}37{:}01]) = 45.4198\ \mathrm{W} \cong 45.4\ \mathrm{W}\,.$$

6.1.2 Determination of PowerTOP’s dynamic overhead

Case 1: logging to hard disk drive

The impact of running PowerTOP while logging to HDD, is shown in Fig. 6. For PowerTOP’s period of activity, the average power is 46.0458 W. While PowerTOP logs to the HDD subsystem, the latter will use power to perform IOPS and will affect concurrent activities on the computer.

Table I shows the power used before, during and after PowerTOP’s measurement and logging activity.

$\overline{p^{(\iota LO)}}[w/o\_ptop] \cong 45.4\text{ W}$ (see Table I), and $\overline{p_{dyn}^{(ptop)}} = 193.4\text{ mW}$

$$\therefore \overline{p^{(\iota LO)}}[with\_ptop] = 46.0458\text{ W}, > \overline{p^{(\iota LO)}}[w/o\_ptop] + \overline{p_{dyn}^{(ptop)}}, = 45.6241\text{ W}$$

Note that despite some double counting of dynamic power in the sum $\overline{p^{(\iota LO)}}[w/o\_ptop] + \overline{p_{dyn}^{(ptop)}}$, this latter sum is still significantly lower than $\overline{p^{(\iota LO)}}[with\_ptop]$. This suggests that PowerTOP does not capture all power users while PowerTOP is logging to hard disk drive.

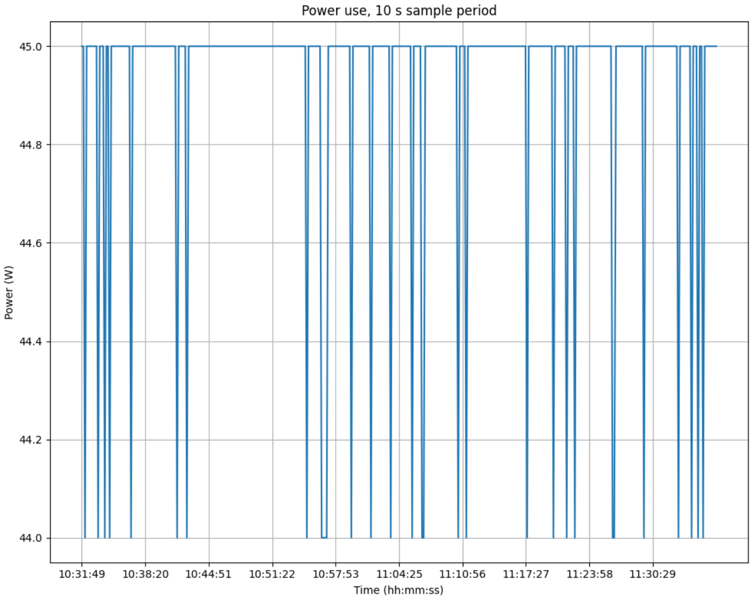


Fig. 5: Power used by the video server's blade, with a quiescent operating system

Table I: Baseline 1, with PowerTOP logging to HDD (Fig. 6)

Power type	**Description**	**Average (W)**
$\overline{p^{(iLO)}}[11{:}11{:}34, 11{:}30{:}02]$	Before starting PowerTOP	45.4189
$\overline{p^{(iLO)}}[11{:}30{:}02, 12{:}02{:}12]$	During PowerTOP's use with HDD	**46.0458**
$\overline{p^{(iLO)}}[12{:}02{:}12, 12{:}17{:}17]$	After PowerTOP' use ended	45.4451
$\overline{p_{dyn}^{(ptop)}}$	Mean dynamic power (PowerTOP)	0.1934
Average power use without PowerTOP, as measured by iLO		45.4307
Average power use without PowerTOP + Mean dynamic power (PowerTOP)		**45.6241**

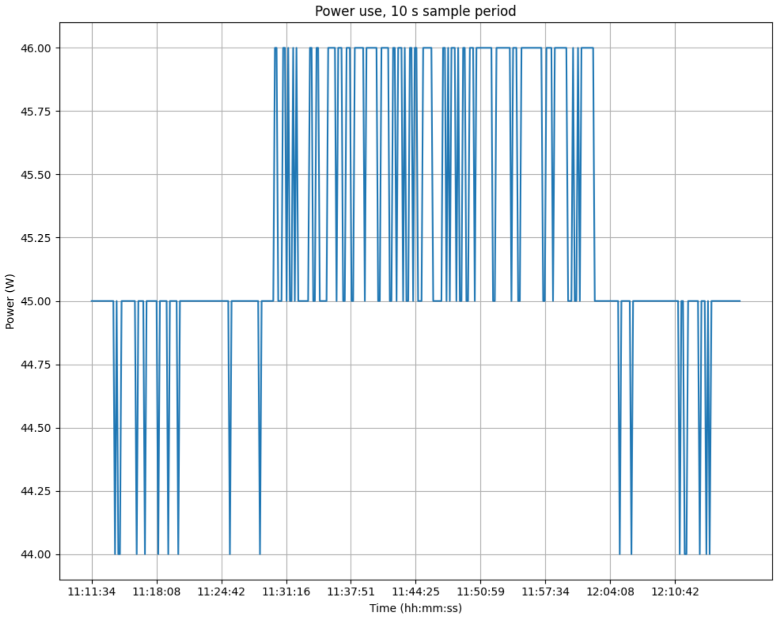


Fig. 6: PowerTOP logging [11:30:02,12:02:12] to HDD has a discernable impact on power use

Case 2: logging to ramdisk

The impact of running PowerTOP while logging to ramdisk, is shown in Fig. 7 and power measurements are shown in Table II.

$$\overline{p^{(\iota LO)}}[w/o\_ptop] \cong 45.4 \text{ W (see Table II).}$$

$$\overline{p_{dyn}^{(ptop)}} = 181.5 \text{ mW}$$

Now: $$\overline{p^{(\iota LO)}}[w/o\_ptop] = P_{idle}^{f_1} + P_{quiescent}^{(OS)}$$

and: $$\overline{p^{(\iota LO)}}[with\_ptop] = P_{idle}^{f_1} + P_{with\_PTOP}^{(OS)} + \overline{p_{dyn}^{(ptop)}}[ptop] \ldots\ldots (3)$$

But: $$\overline{p_{dyn}^{(ptop)}} = P_{with\_PTOP}^{(OS)} + \overline{p_{dyn}^{(ptop)}}[ptop]$$

and: $$P_{with\_PTOP}^{(OS)} \supset P_{quiescent}^{(OS)} \ldots\ldots (4)$$

$$\therefore \overline{p^{(\iota LO)}}[w/o\_ptop] + \overline{p_{dyn}^{(ptop)}} = P_{idle}^{f_1} + P_{quiescent}^{(OS)} + P_{with\_PTOP}^{(OS)} + \overline{p_{dyn}^{(ptop)}}[ptop] \ldots\ldots (5)$$

Relationship (4) is based on the premise that the activities reflected in $P_{quiescent}^{(OS)}$ can reasonably be expected to persist during both the period when PowerTOP is inactive as well as when it is running. Equations (3) and (5) show that the sum of (a) the power measured at the iLO (without PowerTOP running) and (b) the dynamic power measured using PowerTOP, account twice for $P_{quiescent}^{(OS)}$. Indeed, the measurements show a small double accounting:

$$\overline{p^{(\iota LO)}}[w/o\ ptop] + \overline{p_{dyn}^{(ptop)}} = 45.5846\ ,> \overline{p^{(\iota LO)}}[with\_ptop] = 45.5052\ ;$$

$$45.5846 - 45.5052\ \mathrm{W}, = 79.6\ \mathrm{mW}$$

The doubly-counted power consists of $P_{quiescent}^{(OS)}$. Basing on this premise, $P_{quiescent}^{(OS)} \cong 80\ \mathrm{mW}$

Table II: Baseline 1, with PowerTOP logging to ramdisk (Fig. 7)

Power type	**Description**	**Average (W)**
$\overline{p^{(\iota LO)}}[12{:}38{:}14,12{:}57{:}27]$	Before starting PowerTOP	45.4217
$\overline{p^{(\iota LO)}}[12{:}57{:}27,13{:}29{:}30]$	During PowerTOP's use with ramdisk	**45.5052**
$\overline{p^{(\iota LO)}}[13{:}29{:}30,13{:}43{:}26]$	After PowerTOP' use ended	45.3690
$\overline{p_{dyn}^{(ptop)}}$	Mean dynamic power (PowerTOP)	0.1851
Average power use without PowerTOP, as measured by iLO		45.3995
Average power use without PowerTOP + Mean dynamic power (PowerTOP)		**45.5846**

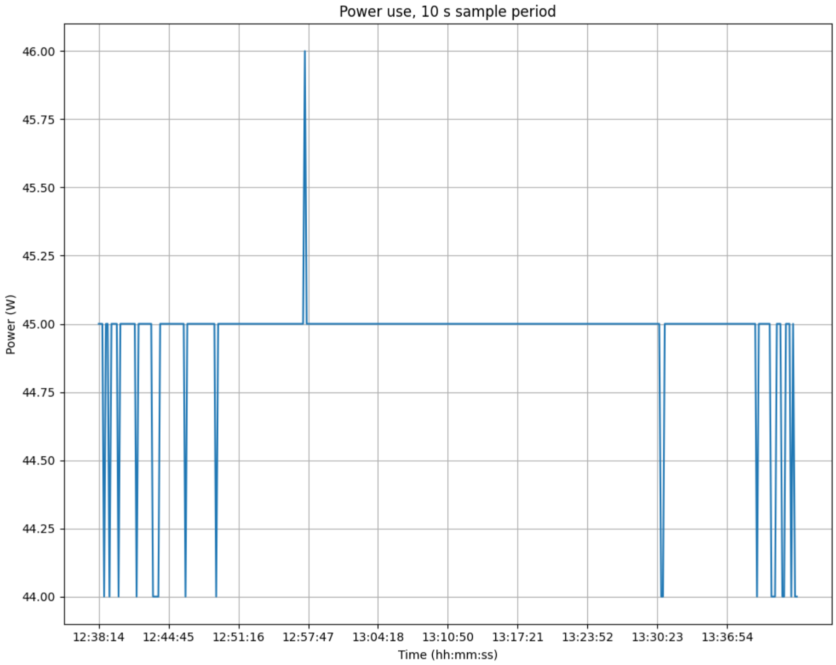


Fig. 7: PowerTOP logging [12:57:27,13:29:30] to ramdisk has a lower impact than logging to HDD.

PowerTOP-process's power use and ranking

Table III shows the processes that cumulatively use 90% of the total power used during PowerTOP's measurement interval.

Table III: Mean power used by processes over measurement period $[12{:}57{:}27,13{:}29{:}30]$

Description	**PW Estimate (mW)**
tick_sched_timer	45.84

toggle_allocation_gate	38.17
fb_flashcursor	19.38
watchdog_timer_fn	15.84
[PID 4085] powertop --csv --time=10 --iteration=180	14.58
handle_update	9.30
bnx2x_period_task	8.23
[PID 119] [kcompactd0]	7.82
bnx2x_sp_task	6.70
pci_pme_list_scan	3.91

Fig. 8 shows the variation of these processes' power use over the period of measurement.

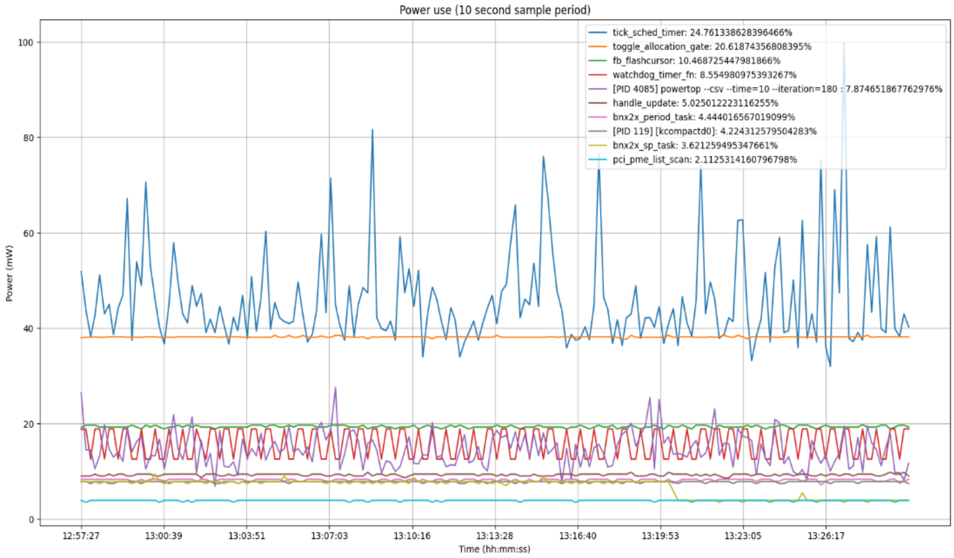


Fig. 8: Processes' power use with time, in descending order, up to 90% of total mean power

6.1.3 Baseline 1: f_1

The measurements reported in

Table I and Table II suggest that $P_{idle}^{f_1} + P_{quiescent}^{(OS)} \cong 45.4$ W. Moreover, Fig. 9 and Fig. 10 show that $f_1 \cong 1.2$ GHz ; taking the mean across all 16 threads reveals an average frequency of 1213 MHz. The top row of each set of graphs duplicates the power measured at the iLO. Each graph in each other row shows frequency against time, for each of the 8 processor cores. Frequency is measured by PowerTOP; hence frequency readings are only available for the time during which PowerTOP is active.

Each core supports two hardware threads; hence the two plots on each graph. In both cases, core 1 (CPU1 and CPU9) is most active during PowerTOP's execution.

6.1.4 First observation: log to ramdisk

It has been seen (case 1 and case 2) that while logging to ramdisk, the sum of $\overline{p^{(\iota LO)}}[without\_powertop]$ and $\overline{p_{dyn}^{(ptop)}}$ is significantly closer to $\overline{p^{(\iota LO)}}[with\_powertop]$, than for the case of logging to HDD. This suggests that PowerTOP better captures dynamic power used by the computer system, when it logs to ramdisk than to HDD. Ramdisk is henceforth used as logging destination. **More generally, it suggests that PowerTOP does not capture dynamic power used by the hard disk drives[76].** This was confirmed through discussion with PowerTOP's maintainers [380].

---

[76] In the case of the hard disk drive, dynamic power refers to the power used to serve disk IOPS.

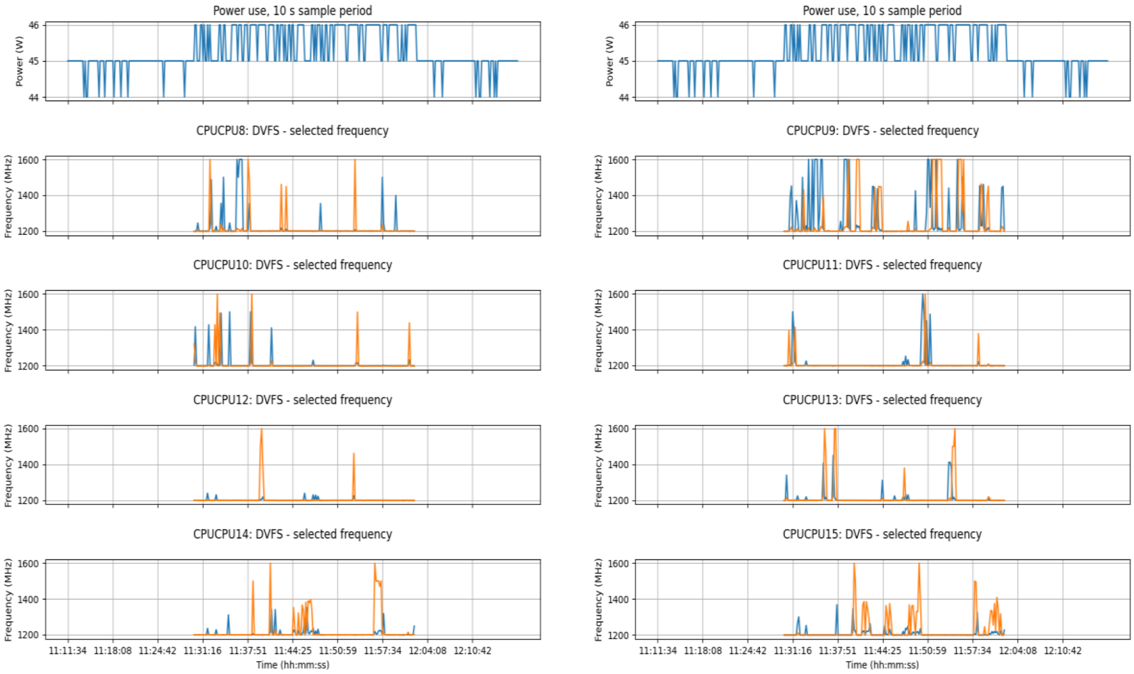


Fig. 9: While logging to HDD, most of the time is spent at 1.2 GHz, with short frequency excursions

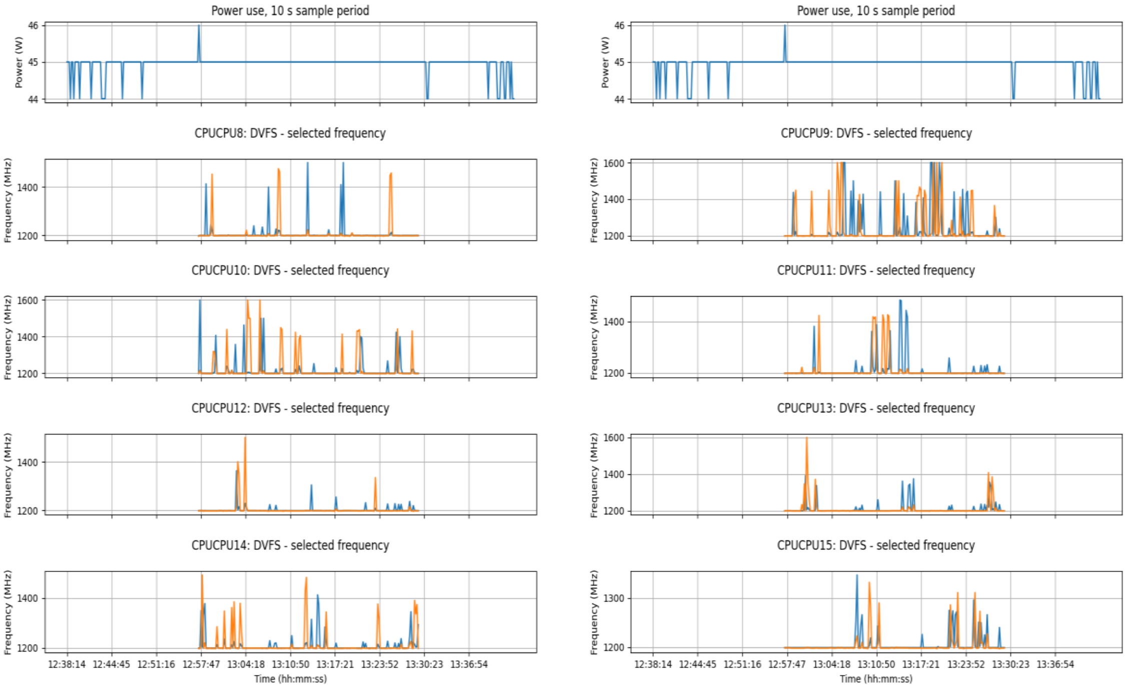


Fig. 10: While logging to ramdisk, most of the time is spent at 1.2 GHz, with short frequency excursions

6.2 Video server's Baseline 2: $P_{b2}^{video}, = P_{idle}^{f_2} + P_{quiescent}^{(OS+dockerd+containerd)}$

6.2.1 A first attempt at baseline 2

Measurements are summarized in Table IV and Fig. 11.

Table IV: Baseline 2 (see Fig. 11)

Power type	Description	Average (W)
$\overline{p^{(\imath LO)}}[12{:}48{:}24,13{:}47{:}53]$	Before starting PowerTOP	45.5028
$\overline{p^{(\imath LO)}}[13{:}47{:}53,14{:}52{:}29]$	During PowerTOP's use with ramdisk	**45.5469**
$\overline{p^{(\imath LO)}}[14{:}52{:}29,15{:}50{:}30]$	After PowerTOP's use ended	45.5
$\overline{p_{dyn}^{(ptop)}}$	Mean dynamic power (PowerTOP)	0.7727
Average power use without PowerTOP, as measured by iLO		45.5

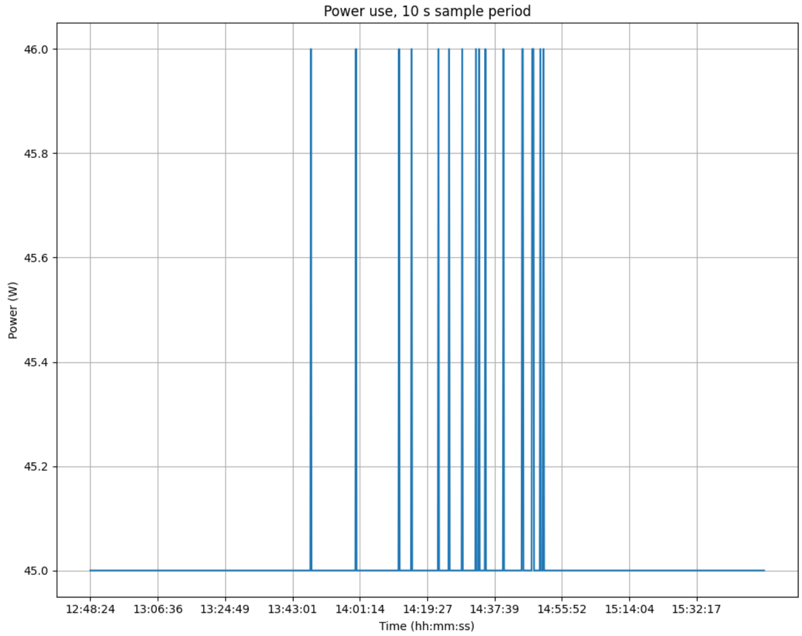


Fig. 11: $P_{idle}^{f_2} + P_{quiescent}^{(OS+dockerd+containerd)}$, as measured by iLO instrumentation

A first glance at Table IV suggests that power use at baseline 2 is only marginally higher than at baseline 1 (45.5 W vs 45.4 W). This impression is supported by the following observations:

1. $f_2 \cong 1.2$ GHz : taking the mean across all 16 threads reveals an average frequency of 1210 MHz, and
2. there is no apparent use of the HDD.

These suggest no significant change in idle power use $\left(P_{idle}^{f_2} \cong P_{idle}^{f_1}\right)$.

On the other hand, the difference between $p_{dyn}^{(ptop)}(baseline\_2)$ and $p_{dyn}^{(ptop)}(baseline\_1)$ is 0.7727 – 0.1851, = 0.5876 W. This is mostly attributable to two Docker containerd processes, which use about 0.48 W. Now $\overline{p^{(\iota LO)}}[12{:}48{:}24,13{:}47{:}53]$ and $\overline{p^{(\iota LO)}}[14{:}52{:}29,15{:}50{:}30]$ are truncated consistently by the iLO to 45 W. This suggests that the dynamic power contributed by Docker components is insufficient to raise the power to 46W in the period $[13{:}47{:}53,14{:}52{:}29]$. There is an evident difference in dynamic power between the two power baselines, and it does not emerge from the use of the iLO's power meter. A different approach is merited here, one in which PowerTOP's dynamic power measurement is used to obtain baseline 2.

6.2.2 A second attempt at baseline 2

The approach referred to in sub-section 5.3 will be used. The difference in dynamic power will be added to the baseline 1 power to obtain the baseline 2 power.

Now:

$$\Delta p_{dyn}^{(ptop)} = p_{dyn}^{(ptop)}(baseline\_2) - p_{dyn}^{(ptop)}(baseline_1) = 0.7727 - 0.1851 = 0.5876\text{ W}$$

But measurements show that (see Table V):

$$\Delta p_{dyn}^{(ptop)}(ptop\_baseline\_2) = 14.3740\text{ mW --- this is the difference due to PowerTOP.}$$

$$\therefore P_{idle}^{f_2} + P_{quiescent}^{(OS+dockerd+containerd)} = P_{idle}^{f_1} + P_{quiescent}^{(OS)} + \Delta p_{dyn}^{(ptop)} - \Delta p_{dyn}^{(ptop)}(ptop)$$

$$= 45.4 + 0.5876 - 0.0144 \cong 45.97\text{ W}$$

This is consistent with the graphical summarization of iLO measurements shown in Fig. 11.

Table V: Mean power used by processes over measurement period $[13{:}47{:}53,14{:}52{:}29]$

Description	PW Estimate (mW)
[PID 3853] containerd --config /var/run/docker/containerd/containerd.toml	239.0722222
tick_sched_timer	126.9066667
[PID 3854] containerd --config /var/run/docker/containerd/containerd.toml	44.86223889
[PID 3871] containerd --config /var/run/docker/containerd/containerd.toml	44.60930833
[PID 3858] containerd --config /var/run/docker/containerd/containerd.toml	41.63066389
toggle_allocation_gate	38.17083333
[PID 3856] containerd --config /var/run/docker/containerd/containerd.toml	36.41606389
[PID 3872] containerd --config /var/run/docker/containerd/containerd.toml	34.91716944
[PID 3855] containerd --config /var/run/docker/containerd/containerd.toml	33.50985
fb_flashcursor	19.36638889
watchdog_timer_fn	15.74722222
[PID 4091] powertop --csv --time=10 --iteration=360	14.37397222
[PID 17] [rcu_preempt]	11.21458333

6.2.3 Reaffirmation of PowerTOP-process’s power use

The measurement shown in Table V reaffirms that shown in Table III. The proximity of the measurements (14.58 and 14.37 mW) supports the essential dimension of reproducibility of scientific experimentation.

6.3 Virtual switch’s Baseline 1: $P_{idle}^{f_1} + P_{quiescent}^{(OS)}$

6.3.1 Average total power

Since the iLO truncates $[n, n+1)$ to n, then the computation of the mean will count the incidences of 45 W and 44 W, and use them as weights to compute a lower limit on the range of values for the ten-second mean value average. Using this premise, the average ten-second mean power measured by the iLO, under the condition of a quiescent operating system (see Fig. 12), is the following:

$$P_{idle}^{f_1} + P_{quiescent}^{(OS)} = \overline{p^{(iLO)}}([21{:}43{:}44, 05{:}29{:}18]) = 46.4293\ \text{W}.$$

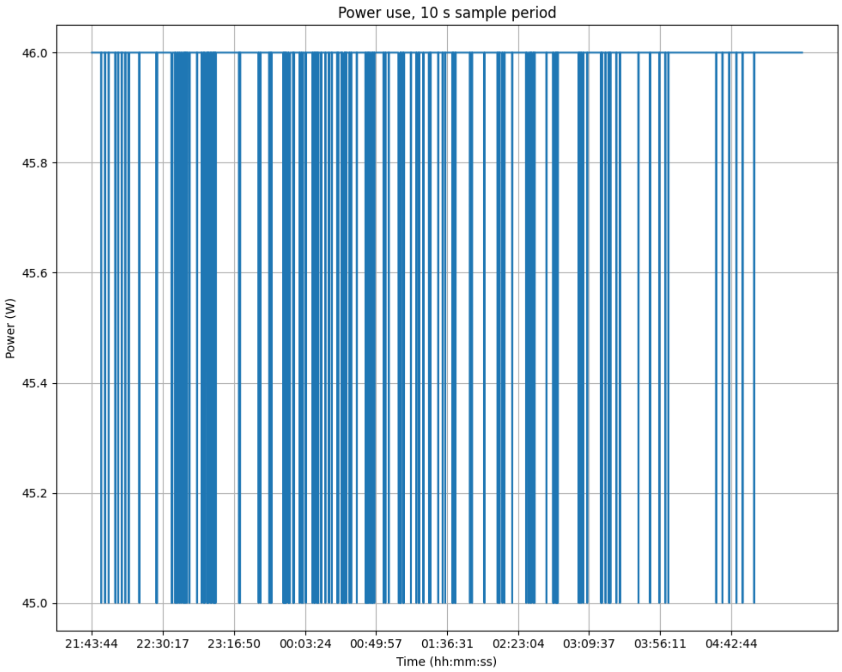


Fig. 12: Power used by the virtual switch’s blade, with a quiescent operating system

6.3.2 Average dynamic power

Dynamic power is measured through the use of PowerTOP. With the virtual switch in the state corresponding to baseline 1, concurrent iLO and PowerTOP measurements are taken. Results are summarized in Fig. 13 and Table VI.

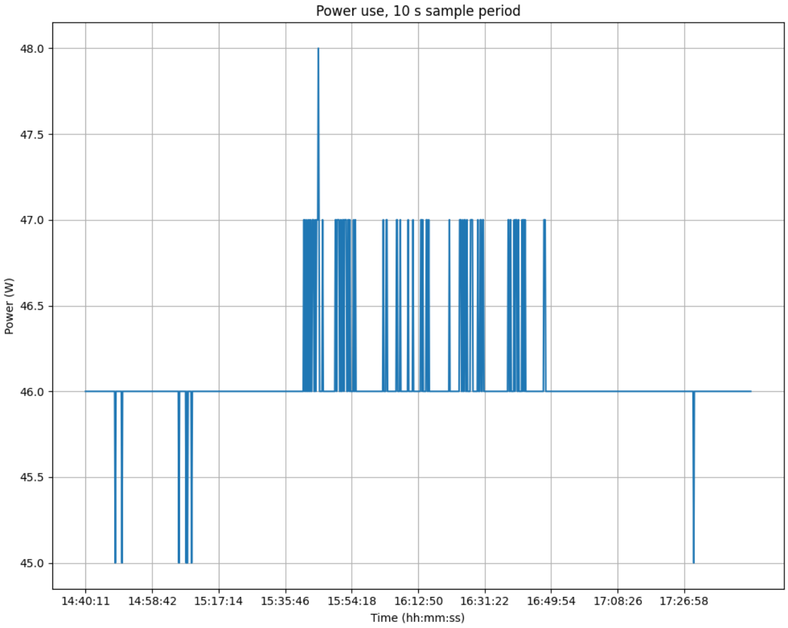


Fig. 13: PowerTOP logging to ramdisk takes place during the period [15:45:00,16:49:31]

Average dynamic power measured is 191.5 mW. Moreover, although average total power use (without PowerTOP running) is about 80 mW higher (46.5098 vs 46.4293 W), the lower value was obtained over a much longer period of time. Therefore, the lower value will be retained.

Table VI

Power type	**Description**	**Average (W)**
$\overline{p^{(iLO)}}[14:40:11,15:45:00]$	Before starting PowerTOP	46.5208
$\overline{p^{(iLO)}}[15:45:00,16:49:31]$	During PowerTOP's use	**46.6797**
$\overline{p^{(iLO)}}[16:49:31,17:45:30]$	After PowerTOP's use ended	46.4970
$\overline{p_{dyn}^{(ptop)}}$	Mean dynamic power (PowerTOP)	0.1915
Average power use without PowerTOP, as measured by iLO		46.5098

6.4 Virtual Switch's Baseline 2: $P_{idle}^{f_2} + P_{quiescent}^{(OS+ovs-vswitchd+ovsdb-server)}$

Measurements are summarized in Table VII and Fig. 14. Furthermore, Fig. 15 shows the processes which consumed the most power, accumulated up to the 90th percentile.

Table VII

Power type	Description	Average (W)
$\overline{p^{(iLO)}}[01{:}54{:}28,03{:}00{:}02]$	Before starting PowerTOP	47.5026
$\overline{p^{(iLO)}}[03{:}00{:}02,04{:}04{:}49]$	During PowerTOP's use	47.5935
$\overline{p^{(iLO)}}[04{:}04{:}49,04{:}59{:}49]$	After PowerTOP's use ended	47.5092
$\overline{p_{dyn}^{(ptop)}}$	Mean dynamic power (PowerTOP)	0.6735
Average power use without PowerTOP, as measured by iLO		47.5056

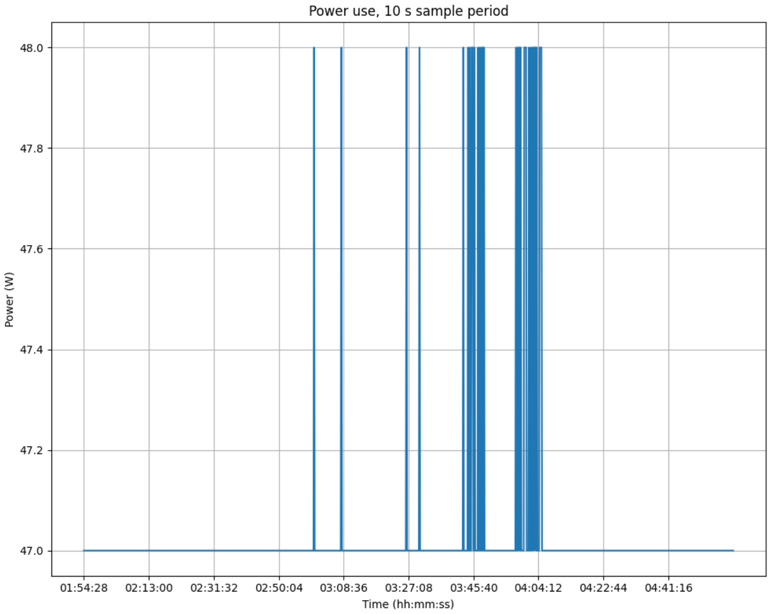


Fig. 14: $P_{idle}^{f_2} + P_{quiescent}^{(OS+ovs-vswitchd+ovsdb-server)}$, as measured by iLO instrumentation

Measurements summarized in Table VII suggest that:

$$P_{idle}^{f_2} + P_{quiescent}^{(OS+ovs-switchd+ovsdb-server)} \cong 47.5\text{ W}$$

6.5 Re-evaluation of the virtual switch's baselines

Before approaching the re-evaluation of the virtual switch's baselines, a review of the re-evaluation applied to the video server's data may be helpful. For the video server, denoting baselines 1 and 2 by P_{b1}^{video} and P_{b2}^{video}respectively,

$$P_{b1}^{video} = P_{idle}^{f_1} + P_{quiescent}^{(OS)} \cong 45.4\text{ W}$$

and:

$$P_{b2}^{video} = P_{b1}^{video} + \Delta p_{dyn}^{(ptop)} - \Delta p_{dyn}^{(ptop)}(ptop) \cong 44.9 + 0.5876 - 0.0144 = 45.97\text{ W}$$

$P_{b2}^{video}$was obtained through the sum of $P_{b1}^{video}$and the increase in dynamic power because the iLO's measurement (Fig. 11) does not account for the observed difference between dynamic power (as measured by PowerTOP) in baselines 1 and 2 respectively.

Now: for the virtual switch, denoting baselines 1 and 2 by P_{b1}^{switch} and P_{b2}^{switch} respectively:

$$P_{b1}^{switch} = P_{idle}^{f_1} + P_{quiescent}^{(OS)} \cong 46.4 \text{ W}$$

and:

$$P_{b2}^{switch} = P_{idle}^{f_2} + P_{quiescent}^{(OS+ovs-vswitchd+ovsdb-server)} \cong 47.5 \text{ W}$$

However, these two baselines are irreconcilable with the difference in dynamic power. Baseline 1's dynamic power is 0.1915 W, whereas baseline 2's is 0.6735 W, i.e., $\Delta p_{dyn}^{(ptop)} = 0.482$ W.

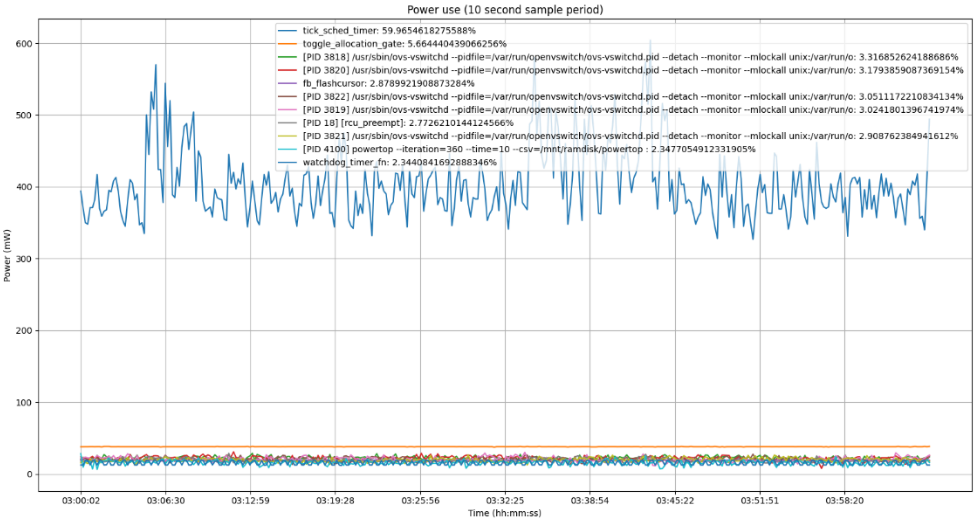


Fig. 15: Processes' power use with time, in descending order, up to 90% of total mean power

In this case, it is useful to consider the full range of possible values which the baselines may take, i.e.:

$P_{b1}^{switch} \in [45.9, 46.9]$,

and: $P_{b2}^{switch} \in [47.0, 48.0]$

With these ranges of possible values and an expected difference between the two actual values of 0.482W, reconciliation between all measurements is possible. If the dynamic power difference between the two baselines is reflected in the estimates P_{b1}^{switch} and P_{b2}^{switch} by withdrawing from the upper and lower range limits in equal parts, then:

$$P_{b1}^{switch} \cong 46.9 - \frac{0.382}{2} \cong 46.7 \text{ W and}$$

$$P_{b2}^{switch} \cong 47.0 + \frac{0.382}{2} \cong 47.2 \text{ W.}$$

6.6 Power use during video service operation

6.6.1 Manual service management

Containerized operation

Results are shown in Table VIII. Fig. 16 and Fig. 17 show PowerTOP's measurements offset by baseline 1 and 2 respectively. Note that the pre-instance average is consistent with the result obtained in sub-sub-section 6.2.2 (baseline 2). The post-instance average is consistent too; a cause of the marginal increase is the residual activity of the container after streaming ends; furthermore, PowerTOP was operated beyond service end, to observe the measured difference in dynamic power between service and post-service operation.

Table VIII

Power type	**Description**	**Average (W)**	**Range (W)**
$\overline{p^{(\iota LO)}}[13:34:23,14:04:14]$	Before starting a service instance	45.51	[45.01,46.01]
$\overline{p^{(\iota LO)}}[14:04:14,15:36:16]$	During the service instance's operation	**46.56**	**[46.06,47.06]**
$\overline{p^{(\iota LO)}}[15:36:16,15:47:50]$	After the service instance ended	46.13	[45.63,46.63]
$\overline{p_{dyn}^{(ptop)}}[14:04:14,15:36:16]$	Mean dynamic power (PowerTOP) during the service instance's operation	2.1295	N/A
$\overline{p_{dyn}^{(ptop)}}[15:36:16,15:53:03]$	Mean dynamic power (PowerTOP) after the service instance's operation	1.2449	N/A

Fig. 17 suggests that baseline 1 is a more accurate reference; this might arise out of double accounting for containerd processes (see Table V). Cross-analysis of the process power use (i.e., comparing power used in baseline 2 with that used in the minimum scaleLevel) can shed further light. Indeed, Table IX reveals that the containerd processes would be doubly counted. Baseline 1 is the better reference.

Table IX reveals two new significant participants in the power use scheme:

- net_rx
    - index number 3 in the list of kernel soft IRQs
- ffmpeg -re -i ./chosen.mp4 -c:v copy -f mpegts tcp://10.0.0.1:7778?listen
    - process ID 4232

ffmpeg's role is evident and the activity of the software interrupt handler for processing of received packets is predictable. **Error! Reference source not found.** is a real-time elaboration of the data s ummarized in Table IX. Given the density of the plot, a lower percentile of 50 is used to improve clarity.

Fig. 19 shows that the effect of video cessation on net_rx is particularly impactful: it drops from the third highest consumer to zero. The increase in power use by containerd, post-operations, seems to be due to the activity of the "`docker rm`" command on the container which has ceased operations and exited.

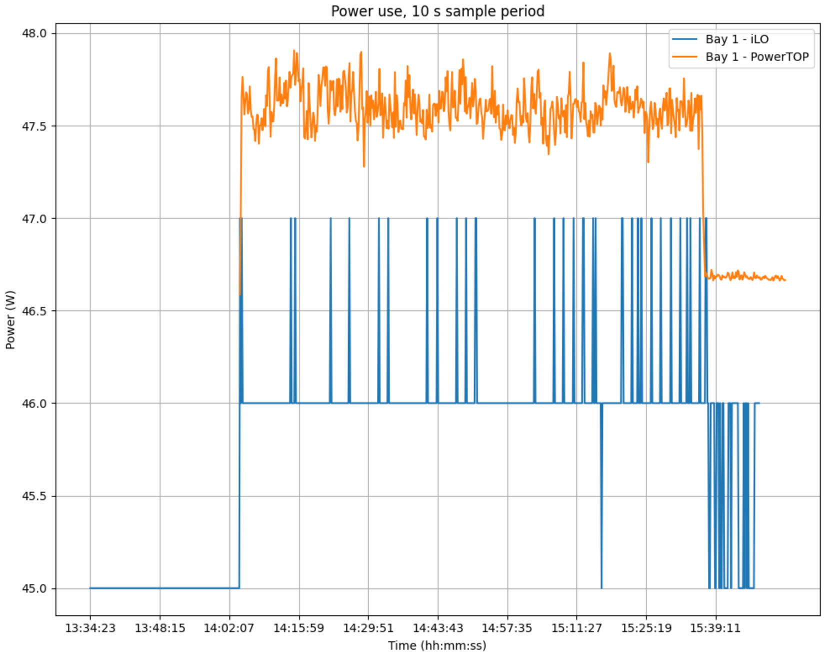


Fig. 16: Video server's power use during containerized service operation. Baseline 2 added to PowerTOP measurements

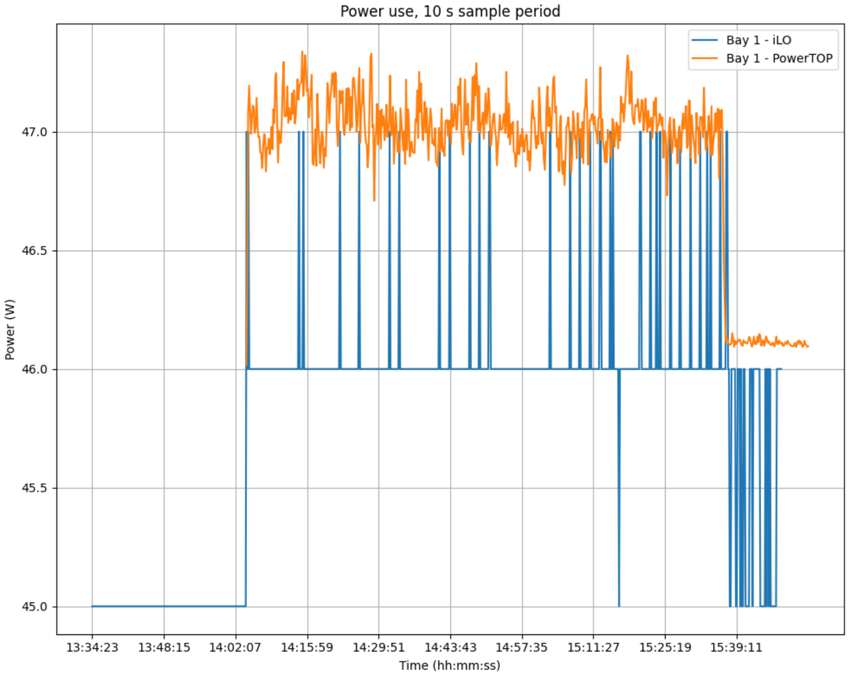


Fig. 17 Video server's power use during containerized service operation. Baseline 1 added to PowerTOP measurements

Table IX: Processes in descending order of mean power use, up to 90th percentile of total

Description	PW Estimate (mW)
tick_sched_timer	453.3066667
[PID 3849] containerd --config /var/run/docker/containerd/containerd.toml	355.3566667
[3] net_rx(softirq)	322.695685
[PID 4199] /usr/bin/containerd-shim-runc-v2 -namespace moby -id 4618343bd39e3412ee6c5ee32fea672f1d0491bf23ecd7cd8b51ce2ee6f1488	111.1032333
[PID 4232] ffmpeg -re -i ./chosen.mp4 -c:v copy -f mpegts tcp://10.0.0.1:7778?listen	107.215
[PID 3859] containerd --config /var/run/docker/containerd/containerd.toml	69.19619333
[PID 3850] containerd --config /var/run/docker/containerd/containerd.toml	68.50957333
[PID 3857] containerd --config /var/run/docker/containerd/containerd.toml	49.95857167
[PID 3862] containerd --config /var/run/docker/containerd/containerd.toml	48.101035

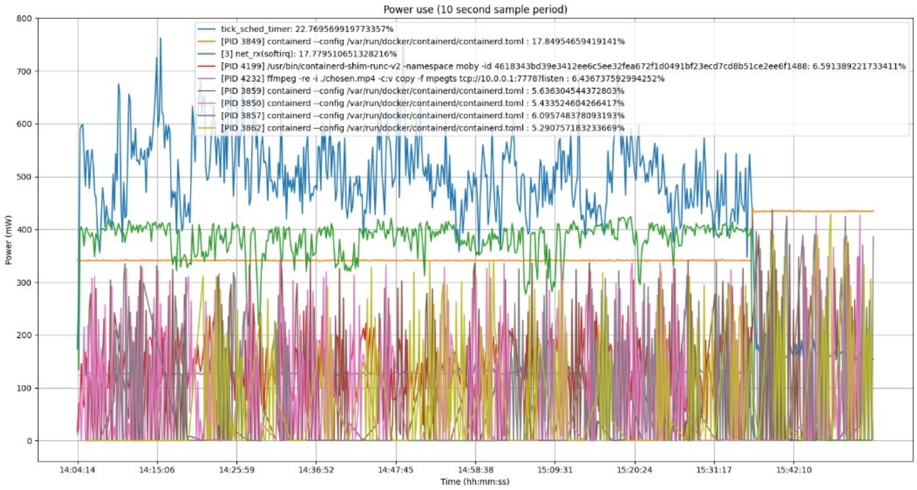


Fig. 18: Process power use, in descending order up to the 90th percentile of total mean power

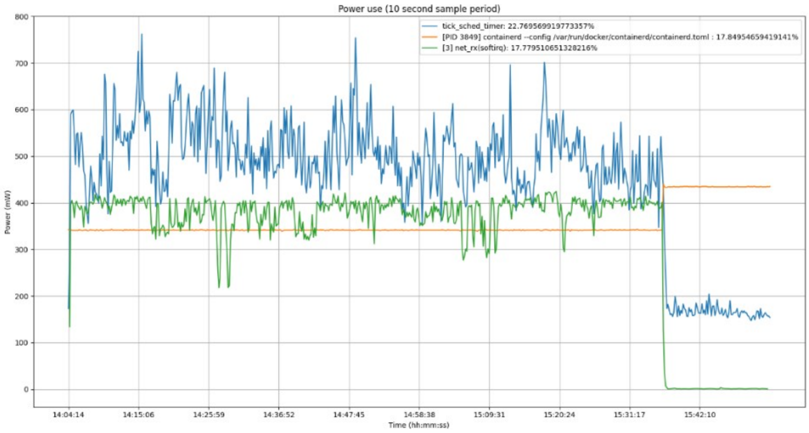


Fig. 19: Process power use, in descending order up to the 50th percentile of total mean power

Native operation

Results are shown in Table X and Fig. 20.

- Table *X* shows that mean power use before service initiation is the same as P_{b1}^{video}, as expected. It also shows that the mean dynamic power use during service delivery is 1.0412 W, which is less than half that used during containerized operation of the service (2.9196 W).
- Fig. 20 shows that P_{b1}^{video} is a very good estimate of the base platform power. When P_{b1}^{video} is added to the mean dynamic power both during ($\overline{p_{dyn}^{(ptop)}}[12:30:12,14:02:14]$), as well as after($\overline{p_{dyn}^{(ptop)}}[14:02:14,14:19:03]$) service instance operation, the sum is a close approximation of the power measured at the iLO.

Table X

Power type	**Description**	**Average (W)**	**Range (W)**
$\overline{p^{(\iota LO)}}[12:24:40,12:30:12]$	Before starting a service instance	45.44	[44.94,45.94]
$\overline{p^{(\iota LO)}}[12:30:12,14:02:14]$	During the service instance's operation	**45.78**	**[45.28,46.28]**
$\overline{p^{(\iota LO)}}[14:02:14,14:19:03]$	After the service instance ended	45.51	[45.01,46.01]
$\overline{p_{dyn}^{(ptop)}}[12:30:12,14:02:14]$	Mean dynamic power (PowerTOP) during the service instance's operation	1.0412	N/A
$\overline{p_{dyn}^{(ptop)}}[14:02:14,14:19:03]$	Mean dynamic power (PowerTOP) after the service instance's operation	0.1777	N/A

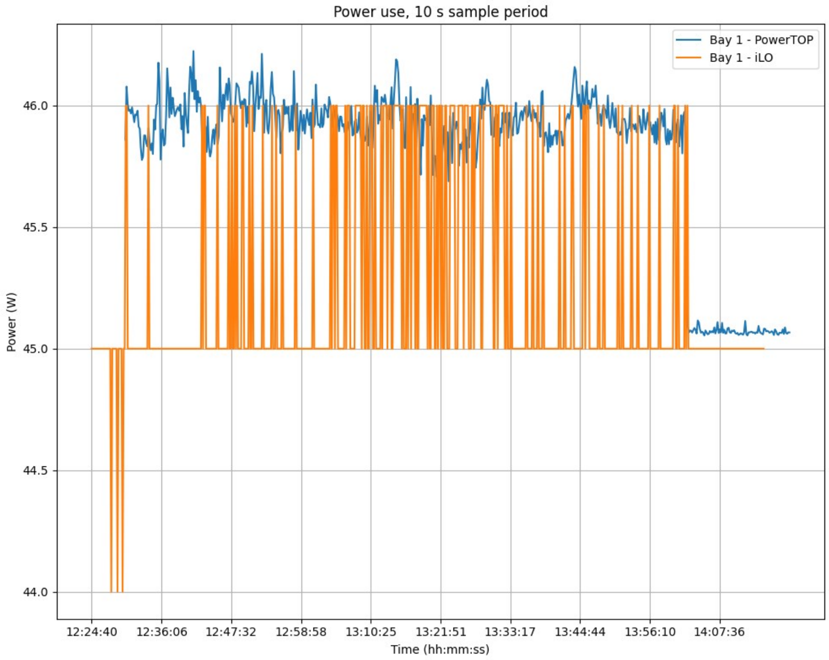


Fig. 20: Video server's power use during native service operation. Baseline 1 added to powertop measurements

6.6.2 Automated service management: Orchestration of containerized streaming

This section presents results collected from running experiments on 1, 2, 5, 10, 20, 40 and 80 instances respectively of the video service. Results consist of:

1. mean power use, both total (iLO instrumentation) and dynamic (PowerTOP instrumentation). Dynamic power data is added to baseline 1 and the sum is plotted on the same Cartesian axes as the total power data.
2. Dynamic power is decomposed into its process power components. The process power components are sorted in descending order of the mean process power over the period of measurement, until some percentage of the mean total dynamic power over the period of measurement, is obtained. Typical percentages are 50 (the 50th percentile) and 80 (80th percentile). The 90th percentile is presented in initial plots, but otherwise avoided, as plots showing the largest power users up to 90% are too dense. Plots and tables are presented to summarize these results.
3. Bitrates of the streams are presented, but since there are as many as 80 streams in the experiments, presentation is delegated to an appendix (Appendix 8). Here, the purpose of presentation is to demonstrate achievement of the QoS KPI, i.e., that the achieved mean bitrate is greater than or equal to the overall mean bitrate of 457 kb/s.

Results for the experiment with the single instance and two instances are accompanied by commentary, to accommodate the reader's grasp of the significance of the results presented. From the 5th instance onward, results are presented in tables and graphs alone. Analysis is delegated to the discussion following the results.

Single instance

Mean power use is shown in Table XI. Fig. 21 shows PowerTOP's measurements offset by baseline 1. The pre-instance average and the post-instance average are very similar to the corresponding statistics obtained with manual service management. The larger post-instance average can be understood after inspecting PowerTOP's measurement of process power use (Fig. 23). Some activity is undertaken by an instance of containerd after the container is destroyed (`docker rm <container name>`), and it persists for some time. However, well after operations end (see Fig. 24), average power use is the same as that found at baseline 2.

Table XI

Power type	**Description**	**Average (W)**	**Range (W)**
$\overline{p^{(iLO)}}[14{:}47{:}05,15{:}03{:}00]$	Before starting a service instance	45.65	[45.15,46.15]
$\overline{p^{(iLO)}}[15{:}03{:}00,15{:}33{:}05]$	During the service instance's operation	**47.03**	**[46.53,47.53]**
$\overline{p^{(iLO)}}[15{:}33{:}05,15{:}52{:}17]$	After the service instance ended	46.17	[45.67,46.67]
$\overline{p_{dyn}^{(ptop)}}[14{:}48{:}17,15{:}03{:}00]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.8593	N/A
$\overline{p_{dyn}^{(ptop)}}[15{:}03{:}00,15{:}33{:}05]$	Mean dynamic power (PowerTOP) during the service instance's operation	1.5940	N/A

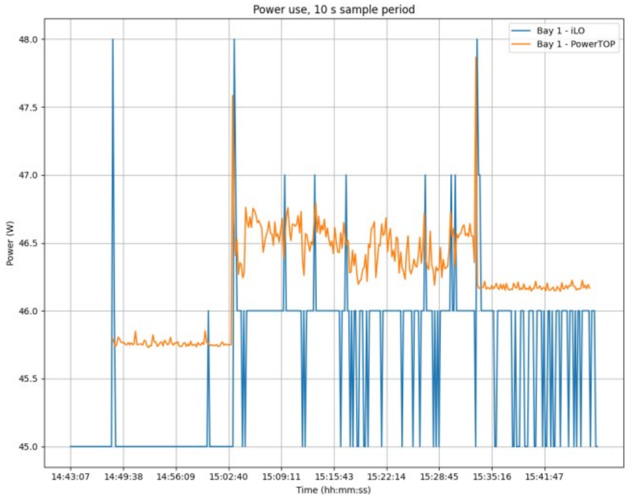


Fig. 21: scaleLevel 1. Video server's power use during containerized service operation. Baseline 1 added to powertop measurements

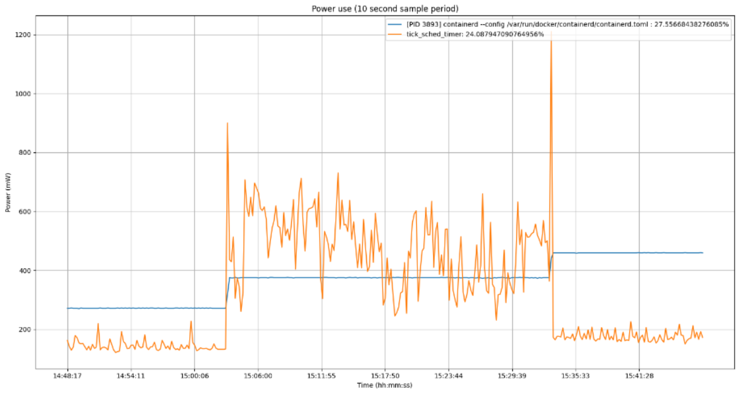


Fig. 22: scaleLevel 1. Process power use during automated containerized service operation, in descending order up to the 50th percentile of total mean power

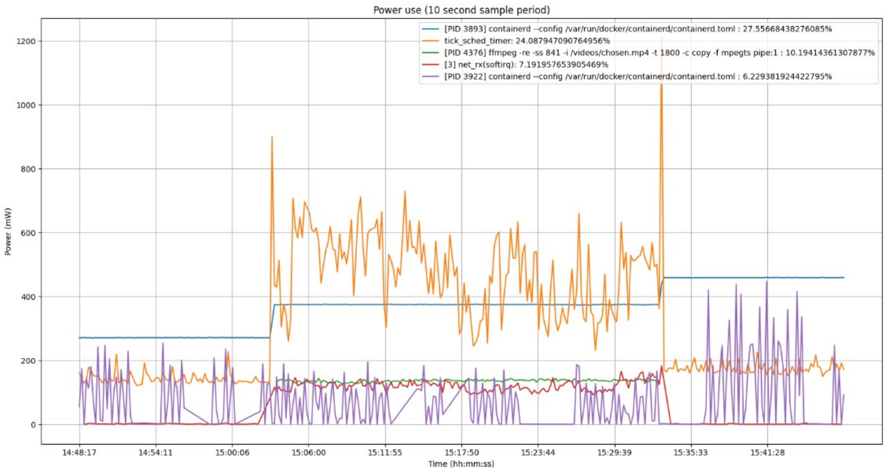


Fig. 23: scaleLevel 1. Process power use during automated containerized service operation, in descending order up to the 75th percentile of total mean power

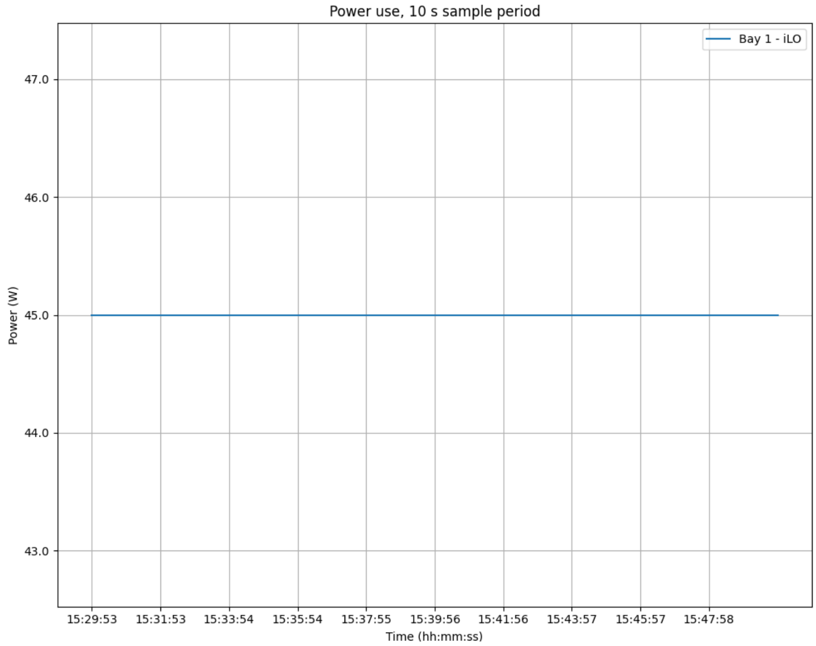


Fig. 24: Well after service operations end, power use returns to baseline 2

Table XII: scaleLevel1: processes in descending order of mean power use, up to 90th percentile of total

Description	**PW Estimate (mW)**
[PID 3893] containerd --config /var/run/docker/containerd/containerd.toml	368.7317
tick_sched_timer	322.3171
[PID 4376] ffmpeg -re -ss 841 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	66.53963
[3] net_rx(softirq)	61.61353
[PID 3922] containerd --config /var/run/docker/containerd/containerd.toml	59.21211
[PID 3920] containerd --config /var/run/docker/containerd/containerd.toml	53.02791
[PID 3898] containerd --config /var/run/docker/containerd/containerd.toml	50.44808
[PID 3894] containerd --config /var/run/docker/containerd/containerd.toml	40.07022
[PID 3907] containerd --config /var/run/docker/containerd/containerd.toml	39.9233
toggle_allocation_gate	38.15671
[PID 18] [rcu_preempt]	32.94561
[PID 3906] containerd --config /var/run/docker/containerd/containerd.toml	19.9275
fb_flashcursor	19.3436
[PID 3923] containerd --config /var/run/docker/containerd/containerd.toml	17.955
[PID 4218] powertop --csv --iteration=328 --time=10	16.48095

Two instances

Mean power use is shown in Table XIII. Fig. 25 shows PowerTOP's measurements laid over the iLO's measurement; offset is baseline 1. Process power use is shown in Fig. 26 (50th percentile) and Fig. 27 (80th percentile). Comparison of Table XIV and Table IX shows that most power is used by the same processes.

As regards service instance dynamic power use, two characterizations can be considered. The first is the actual power use; the second is the differential power use, where "differential" regards the dynamic power use before and during instance operation. Let N service instances be referred to as Ni, for brevity. Comparison of data in Table XIII with data in Table XI reveals:

- actual dynamic power use: 1i vs 2i: 1.5940 vs 2.6162 W
- differential dynamic power use: 1i vs 21: (1.594-0.8593), = 0.7347 vs (2.6162-0.9693) = 1.6469.

Table XIII

Power type	Description	Average (W)	Range (W)
$\overline{p^{(\iota LO)}}[09{:}24{:}01,09{:}44{:}27]$	Before starting a service instance	45.6	[45.10,46.10]
$\overline{p^{(\iota LO)}}[09{:}44{:}27,10{:}14{:}37]$	During the service instance's operation	**47.06**	**[46.56,47.56]**
$\overline{p^{(\iota LO)}}[10{:}14{:}37,10{:}29{:}23]$	After the service instance ended	46.12	[45.62,46.62]
$\overline{p_{dyn}^{(ptop)}}[09{:}29{:}27,09{:}44{:}27]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.9693	N/A
$\overline{p_{dyn}^{(ptop)}}[09{:}44{:}27,10{:}14{:}37]$	Mean dynamic power (PowerTOP) during the service instance's operation	2.6162	N/A

Table XIV: scaleLevel2: processes in descending order of mean power use, up to 90th percentile of total

Description	PW Estimate (mW)
tick_sched_timer	511.7164634
[PID 3882] containerd --config /var/run/docker/containerd/containerd.toml	427.6554878
[3] net_rx(softirq)	114.6671677
[PID 3890] containerd --config /var/run/docker/containerd/containerd.toml	69.27329573
[PID 4440] ffmpeg -re -ss 2211 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	68.02743902
[PID 4446] ffmpeg -re -ss 801 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	67.17682927
[PID 3888] containerd --config /var/run/docker/containerd/containerd.toml	66.81309756
[PID 3884] containerd --config /var/run/docker/containerd/containerd.toml	64.69728354
[PID 18] [rcu_preempt]	58.79012195
[PID 3887] containerd --config /var/run/docker/containerd/containerd.toml	54.53221341
[PID 3895] containerd --config /var/run/docker/containerd/containerd.toml	42.69616463
[PID 3897] containerd --config /var/run/docker/containerd/containerd.toml	38.56689329
toggle_allocation_gate	38.15182927
[PID 3898] containerd --config /var/run/docker/containerd/containerd.toml	32.93696037
[PID 3900] containerd --config /var/run/docker/containerd/containerd.toml	22.35445427
[PID 3885] containerd --config /var/run/docker/containerd/containerd.toml	21.61723171
fb_flashcursor	19.36737805
[PID 4172] powertop --csv --iteration=328 --time=10	17.34588415

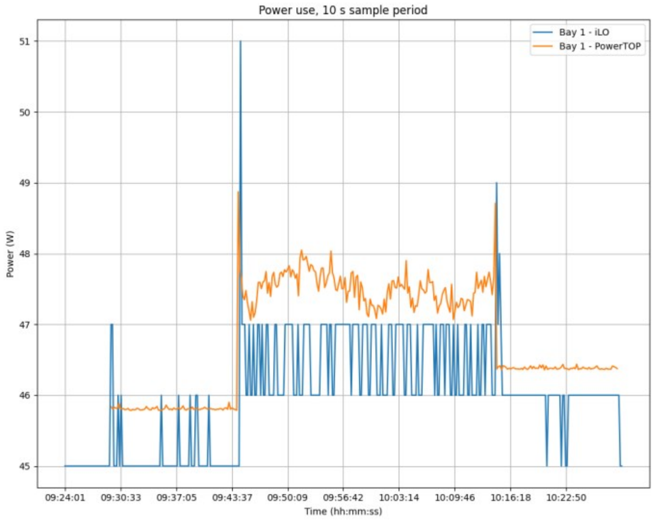


Fig. 25: scaleLevel 2. Video server's power use during containerized service operation.

Fig. 26: scaleLevel 2. Process power use during automated containerized service operation, 50th percentile

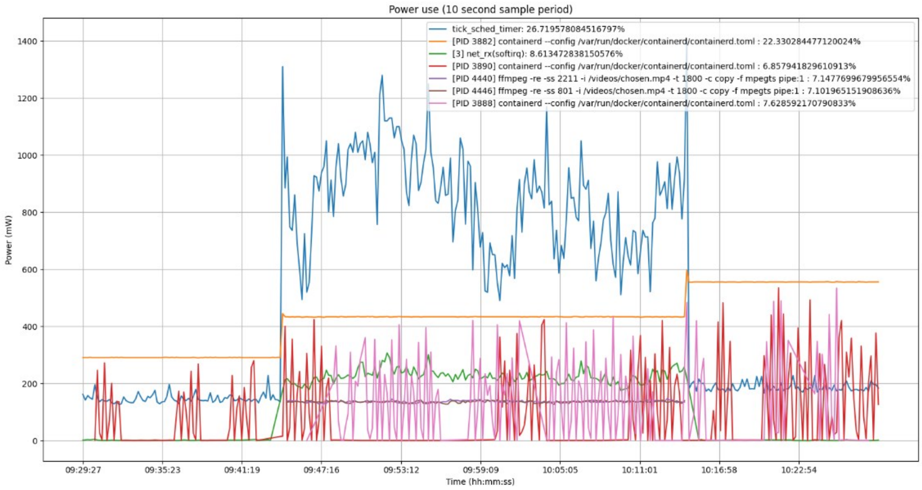


Fig. 27: scaleLevel 2. Process power use during automated containerized service operation, 80th percentile

Five instances

Table XV

Power type	Description	Average (W)	Range (W)
$\overline{p^{(\imath LO)}}[11:29:41,11:49:59]$	Before starting a service instance	45.79	[45.29,46.29]
$\overline{p^{(\imath LO)}}[11:49:59,12:20:15]$	During the service instance’s operation	**48.16**	**[47.66,48.66]**
$\overline{p_{dyn}^{(ptop)}}[11:35:00,11:49:59]$	Mean dynamic power (PowerTOP) before the service instance’s operation	0.9970	N/A
$\overline{p_{dyn}^{(ptop)}}[11:49:59,12:20:15]$	Mean dynamic power (PowerTOP) during the service instance’s operation	4.7421	N/A

Table XVI: scaleLevel5: processes in descending order of mean power use, up to 90th percentile of total

Description	**PW Estimate (mW)**
tick_sched_timer	1123.585366
[PID 3884] containerd --config /var/run/docker/containerd/containerd.toml	479.4939024
[3] net_rx(softirq)	249.058503
[PID 17] [rcu_preempt]	99.97237805
[PID 3886] containerd --config /var/run/docker/containerd/containerd.toml	73.30891463
[PID 4810] ffmpeg -re -ss 1950 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	67.32317073
[PID 4808] ffmpeg -re -ss 1629 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	66.5304878
[PID 4806] ffmpeg -re -ss 2297 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	66.23780488
[PID 4804] ffmpeg -re -ss 1746 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	66.13109756
[PID 4812] ffmpeg -re -ss 2792 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	65.24390244
[PID 4098] containerd --config /var/run/docker/containerd/containerd.toml	60.38504268
[PID 3895] containerd --config /var/run/docker/containerd/containerd.toml	59.62064634
[PID 3889] containerd --config /var/run/docker/containerd/containerd.toml	49.85242988
[PID 3894] containerd --config /var/run/docker/containerd/containerd.toml	47.32973171
[PID 3893] containerd --config /var/run/docker/containerd/containerd.toml	45.44128659
[PID 3887] containerd --config /var/run/docker/containerd/containerd.toml	44.97468902
toggle_allocation_gate	38.13079268
[PID 3897] containerd --config /var/run/docker/containerd/containerd.toml	34.78199695
[PID 3815] /usr/bin/dockerd	28.23655488
[PID 5321] containerd --config /var/run/docker/containerd/containerd.toml	20.61327744
fb_flashcursor	19.35121951

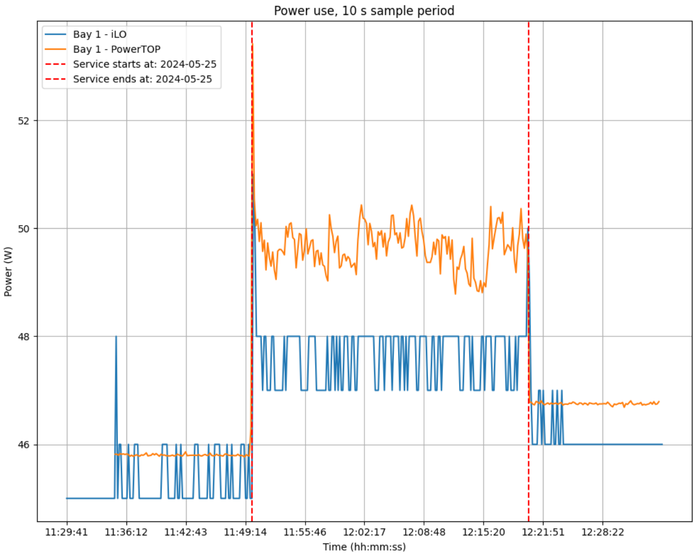


Fig. 28: scaleLevel 5. Video server’s power use during containerized service operation.

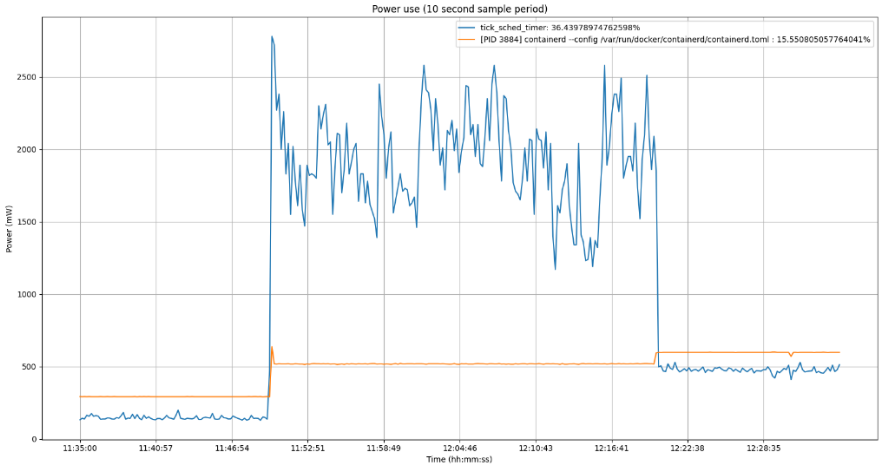


Fig. 29: scaleLevel 5. Process power use during automated containerized service operation, 50th percentile

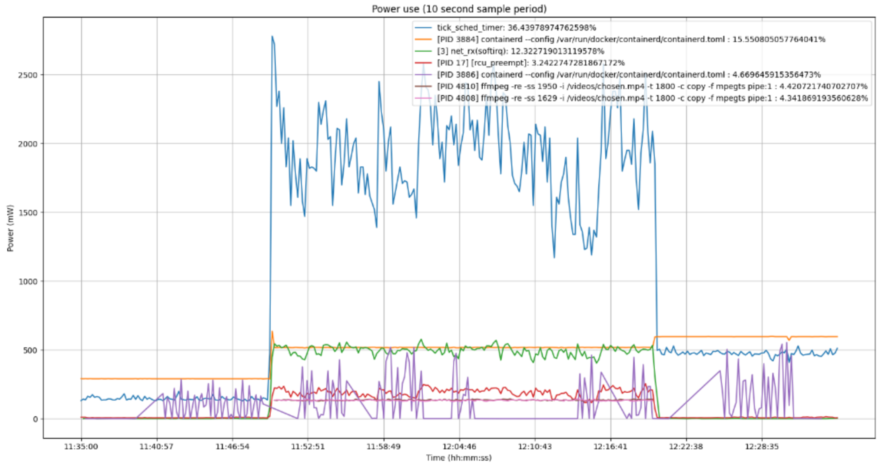


Fig. 30: scaleLevel 5. Process power use during automated containerized service operation, 80th percentile

Ten instances

Table XVII

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[15{:}06{:}49,15{:}27{:}03]$	Before starting a service instance	45.60	[45.10,46.10]
$\overline{p^{(iLO)}}[15{:}27{:}03,15{:}57{:}27]$	During the service instance’s operation	**49.60**	**[49.10,50.10]**
$\overline{p_{dyn}^{(ptop)}}[15{:}12{:}03,15{:}27{:}03]$	Mean dynamic power (PowerTOP) before the service instance’s operation	0.8759	N/A
$\overline{p_{dyn}^{(ptop)}}[15{:}27{:}03,15{:}57{:}27]$	Mean dynamic power (PowerTOP) during the service instance’s operation	8.9781	N/A

Table XVIII: scaleLevel10: processes in descending order of mean power use, up to 90th percentile of total

Description	PW Estimate (mW)
tick_sched_timer	2340.382927
[3] net_rx(softirq)	528.7733811
[PID 3885] containerd --config /var/run/docker/containerd/containerd.toml	524.2865854
[PID 17] [rcu_preempt]	164.6610671
[PID 3896] containerd --config /var/run/docker/containerd/containerd.toml	83.1961372
[PID 3899] containerd --config /var/run/docker/containerd/containerd.toml	76.19127439
[PID 4102] containerd --config /var/run/docker/containerd/containerd.toml	72.0392622
[PID 5403] ffmpeg -re -ss 589 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	65.28353659
[PID 5400] ffmpeg -re -ss 202 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	65.13719512
[PID 5407] ffmpeg -re -ss 2453 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	64.90853659
[PID 5411] ffmpeg -re -ss 121 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	64.74390244
[PID 5395] ffmpeg -re -ss 1012 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	64.57621951
[PID 5401] ffmpeg -re -ss 235 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	64.47865854
[PID 5397] ffmpeg -re -ss 2679 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	64.41768293
[PID 5405] ffmpeg -re -ss 1856 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	64.37195122
[PID 5409] ffmpeg -re -ss 2918 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	63.81097561
[PID 5393] ffmpeg -re -ss 1197 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	63.04878049
[PID 3798] /usr/bin/dockerd	49.08164634
[PID 3895] containerd --config /var/run/docker/containerd/containerd.toml	46.71573171
[PID 5415] containerd --config /var/run/docker/containerd/containerd.toml	43.28141463
[PID 3890] containerd --config /var/run/docker/containerd/containerd.toml	41.41569207
[PID 3892] containerd --config /var/run/docker/containerd/containerd.toml	39.88000915

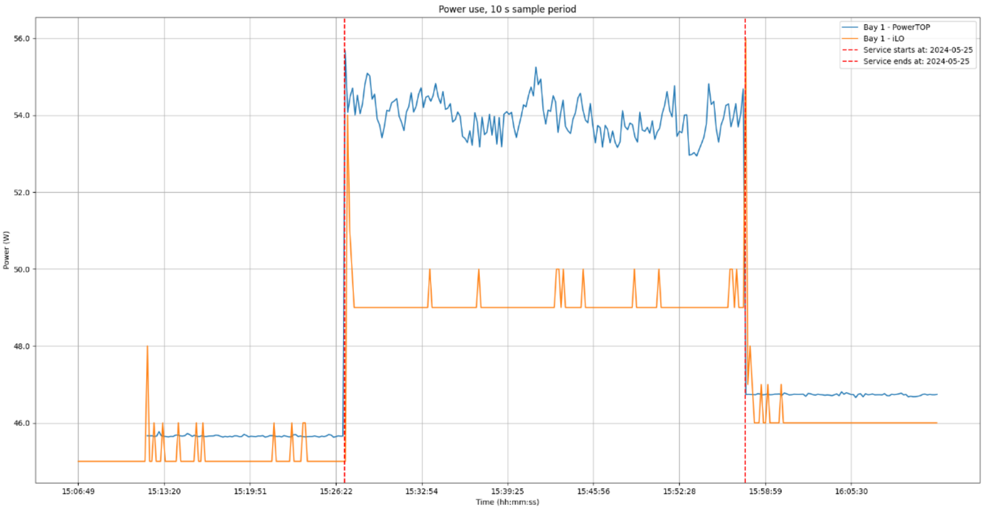


Fig. 31: scaleLevel 10. Video server’s power use during containerized service operation.

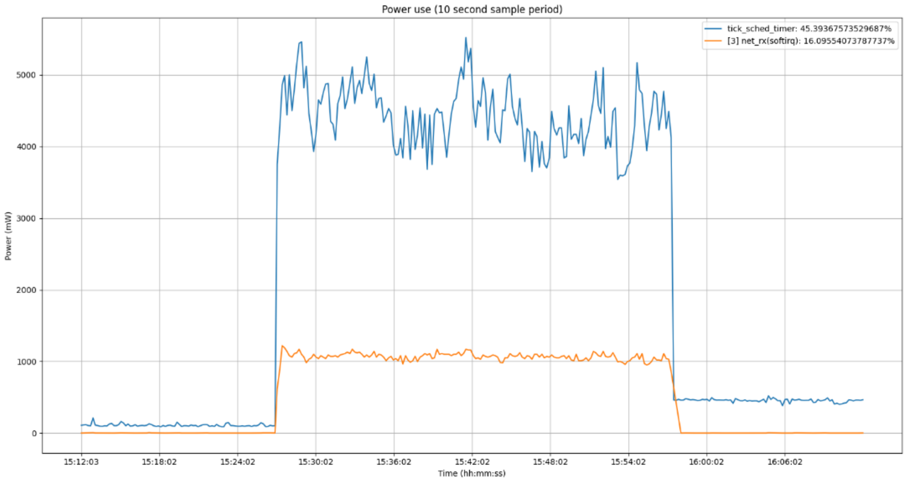


Fig. 32: scaleLevel 10. Process power use during automated containerized service operation, 50th percentile

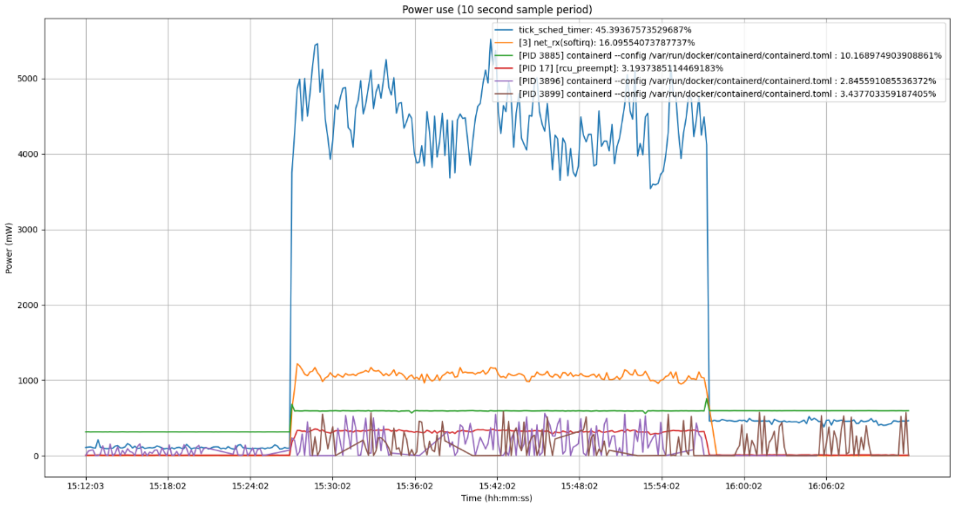


Fig. 33: scaleLevel 10. Process power use during automated containerized service operation, 80th percentile

Twenty instances

Table XIX

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[17{:}56{:}58,18{:}17{:}08]$	Before starting a service instance	45.76	[45.26,46.26]
$\overline{p^{(iLO)}}[18{:}17{:}08,18{:}48{:}00]$	During the service instance’s operation	**51.24**	**[50.74,51.74]**
$\overline{p_{dyn}^{(ptop)}}[18{:}02{:}08,18{:}17{:}08]$	Mean dynamic power (PowerTOP) before the service instance’s operation	0.8913	N/A
$\overline{p_{dyn}^{(ptop)}}[18{:}17{:}08,18{:}48{:}00]$	Mean dynamic power (PowerTOP) during the service instance’s operation	15.3720	N/A

Table XX: scaleLevel20: processes in descending order of mean power use, up to 90th percentile of total

Description	PW Estimate (mW)
tick_sched_timer	4230.064024
[3] net_rx(softirq)	980.6319939
[PID 3882] containerd --config /var/run/docker/containerd/containerd.toml	526.4176829
[PID 17] [rcu_preempt]	171.7254573
hrtimer_wakeup	89.08588415
[PID 3793] /usr/bin/dockerd	85.49856707
[PID 3897] containerd --config /var/run/docker/containerd/containerd.toml	72.28831707
[PID 6622] ffmpeg -re -ss 2972 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	61.27926829
[PID 6542] ffmpeg -re -ss 2862 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.93993902
[PID 6592] ffmpeg -re -ss 893 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.90060976
[PID 6586] ffmpeg -re -ss 990 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.67957317
[PID 6550] ffmpeg -re -ss 1550 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.6179878
[PID 6562] ffmpeg -re -ss 99 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.57012195
[PID 6616] ffmpeg -re -ss 924 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.52530488
[PID 6610] ffmpeg -re -ss 2989 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.52317073
[PID 6634] ffmpeg -re -ss 88 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.4929878
[PID 6602] ffmpeg -re -ss 2953 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.33871951
[PID 6628] ffmpeg -re -ss 1952 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.32164634
[PID 6556] ffmpeg -re -ss 2770 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.31859756
[PID 6520] ffmpeg -re -ss 758 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.27408537
[PID 6580] ffmpeg -re -ss 2899 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.2195122
[PID 6574] ffmpeg -re -ss 504 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	60.03689024
[PID 6546] ffmpeg -re -ss 119 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	59.74237805
[PID 6544] ffmpeg -re -ss 348 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	59.63262195
[PID 6572] ffmpeg -re -ss 563 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	59.43506098
[PID 6548] ffmpeg -re -ss 2838 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	59.41371951
[PID 3885] containerd --config /var/run/docker/containerd/containerd.toml	59.03708232
[PID 6604] ffmpeg -re -ss 1388 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	58.6777439
[PID 3887] containerd --config /var/run/docker/containerd/containerd.toml	58.36759146
[PID 4103] containerd --config /var/run/docker/containerd/containerd.toml	55.89742073
[PID 3896] containerd --config /var/run/docker/containerd/containerd.toml	51.9797561

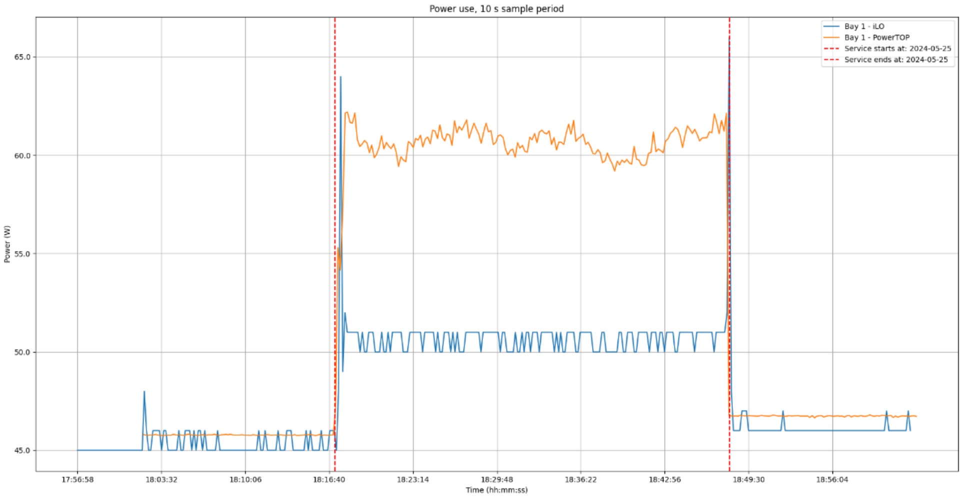


Fig. 34: scaleLevel 20. Video server's power use during containerized service operation.

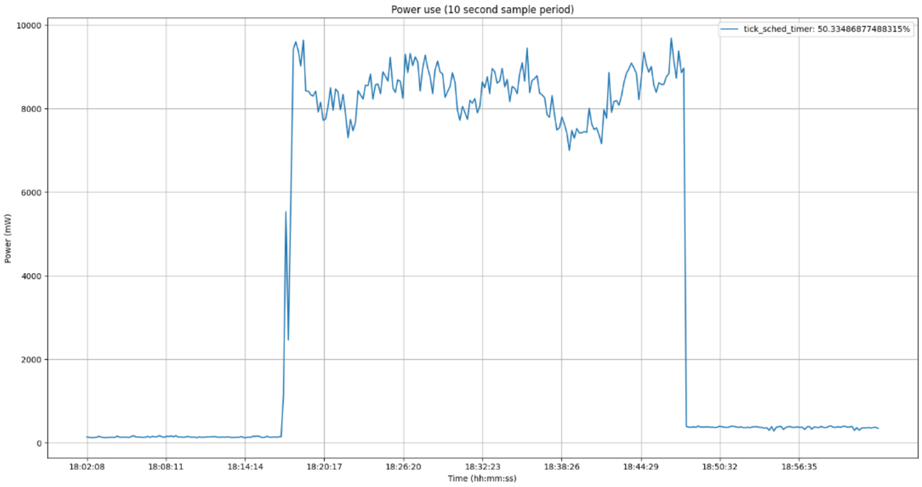


Fig. 35: scaleLevel 20. Process power use during automated containerized service operation, 50th percentile

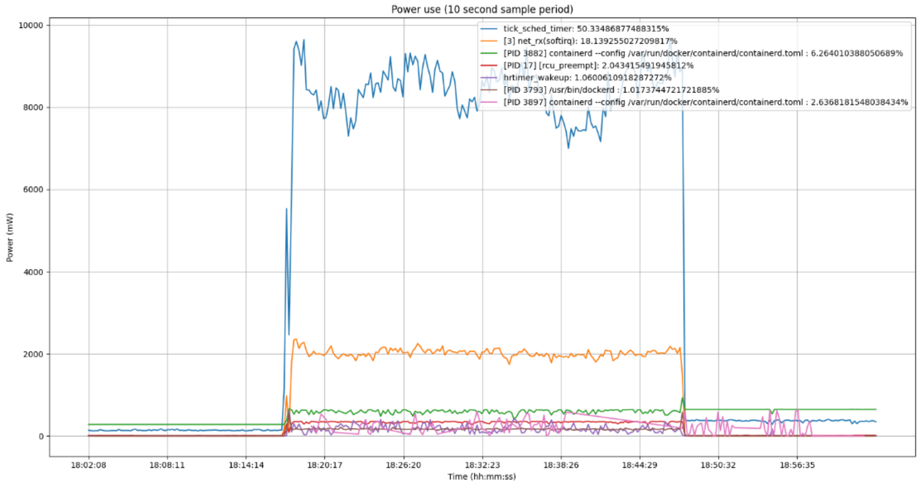


Fig. 36: scaleLevel 20. Process power use during automated containerized service operation, 80th percentile

Forty instances

Table XXI

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[12:57:37,13:17:40]$	Before starting a service instance	45.56	[45.06,46.06]
$\overline{p^{(iLO)}}[13:17:40,13:49:15]$	During the service instance's operation	**53.40**	**[52.90,53.90]**
$\overline{p_{dyn}^{(ptop)}}[13:02:42,13:17:40]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.7206	N/A
$\overline{p_{dyn}^{(ptop)}}[13:17:40,13:49:15]$	Mean dynamic power (PowerTOP) during the service instance's operation	19.1873	N/A

Table XXII: scaleLevel40: processes in descending order of mean power use, up to 90th percentile of total

Description	PW Estimate (mW)
tick_sched_timer	5402.629
[3] net_rx(softirq)	1265.363
hrtimer_wakeup	663.7275
[PID 3882] containerd --config /var/run/docker/containerd/containerd.toml	405.7409
[PID 18] [rcu_preempt]	104.8521
[PID 3795] /usr/bin/dockerd	100.9495
[PID 3899] containerd --config /var/run/docker/containerd/containerd.toml	44.11719
[PID 3884] containerd --config /var/run/docker/containerd/containerd.toml	42.31237
[PID 3887] containerd --config /var/run/docker/containerd/containerd.toml	40.80812
[PID 7927] ffmpeg -re -ss 1413 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	36.59848
[PID 7969] ffmpeg -re -ss 237 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	36.52165
[PID 3898] containerd --config /var/run/docker/containerd/containerd.toml	36.22431
[PID 7933] ffmpeg -re -ss 60 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	36.21494
[PID 7993] ffmpeg -re -ss 2988 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	36.14573
[PID 7913] ffmpeg -re -ss 1714 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	36.06738
[PID 8019] ffmpeg -re -ss 1433 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	36.06707
[PID 8025] ffmpeg -re -ss 805 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.97957
[PID 8005] ffmpeg -re -ss 2716 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.79634
[PID 7833] ffmpeg -re -ss 651 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.77104
[PID 7837] ffmpeg -re -ss 2011 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.70854
[PID 7987] ffmpeg -re -ss 1107 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.70555
[PID 7955] ffmpeg -re -ss 601 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.67683
[PID 8049] ffmpeg -re -ss 2734 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.66189
[PID 7903] ffmpeg -re -ss 1612 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.62622
[PID 7836] ffmpeg -re -ss 1285 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.50274
[PID 7963] ffmpeg -re -ss 765 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.50152
[PID 7867] ffmpeg -re -ss 2176 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.48476
[PID 7999] ffmpeg -re -ss 1579 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.41585
[PID 8012] ffmpeg -re -ss 451 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.3689
[PID 7957] ffmpeg -re -ss 521 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.36555
[PID 7861] ffmpeg -re -ss 1640 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.33659

Description (continued)	**PW Estimate (mW)**
[PID 7843] ffmpeg -re -ss 1723 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.30915
[PID 7953] ffmpeg -re -ss 2340 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.30884
[PID 7879] ffmpeg -re -ss 1688 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.27774
[PID 7889] ffmpeg -re -ss 2893 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.23902
[PID 8057] ffmpeg -re -ss 1396 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.10183
[PID 7901] ffmpeg -re -ss 2755 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.04024
[PID 7849] ffmpeg -re -ss 155 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.03902
[PID 7859] ffmpeg -re -ss 2465 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	35.02957
[PID 8048] ffmpeg -re -ss 351 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	34.9628
[PID 7831] ffmpeg -re -ss 603 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	34.92226
[PID 7919] ffmpeg -re -ss 861 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	34.90823
[PID 7877] ffmpeg -re -ss 1023 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	34.87622

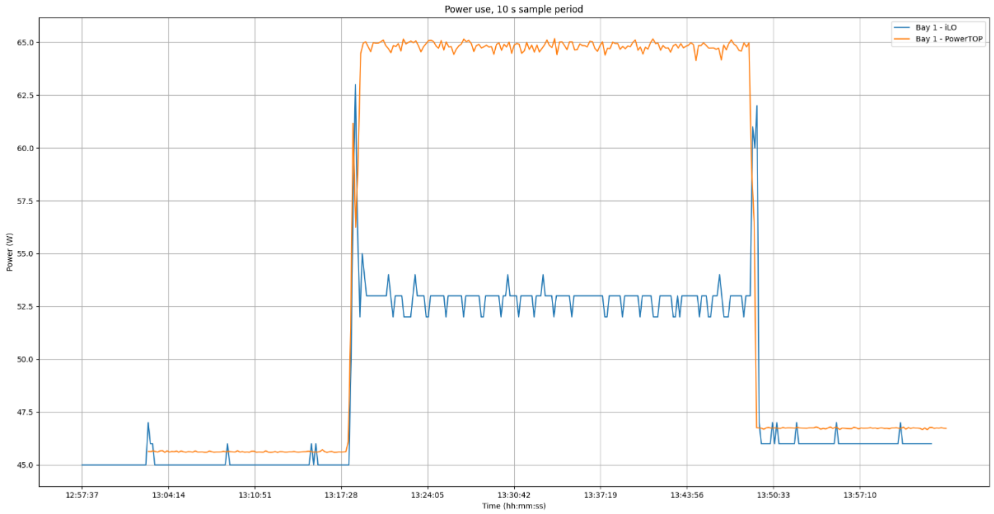


Fig. 37: scaleLevel 40. Video server's power use during containerized service operation.

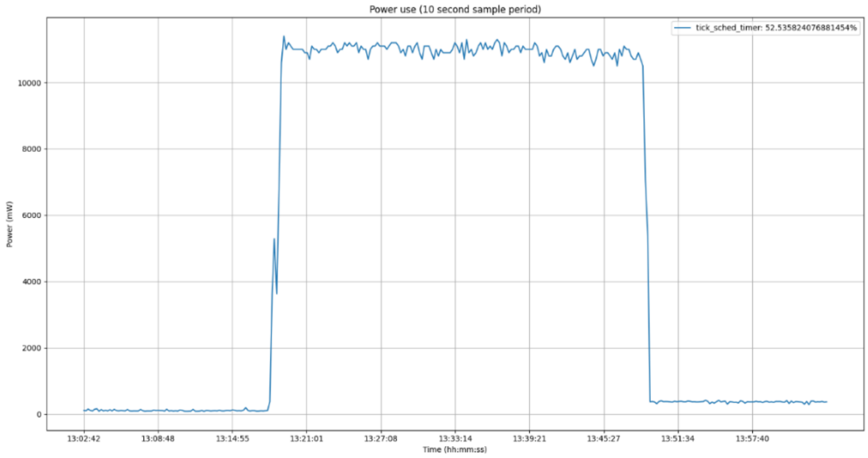


Fig. 38: scaleLevel 40. Process power use during automated containerized service operation, 50th percentile

Fig. 39: scaleLevel 40. Process power use during automated containerized service operation, 80th percentile

Eighty instances

Table XXIII

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[18:19:21, 18:39:20]$	Before starting a service instance	45.53	[45.03,46.03]
$\overline{p^{(iLO)}}[18:39:30, 19:13:53]$	During the service instance's operation	**56.30**	**[55.80,56.80]**
$\overline{p_{dyn}^{(ptop)}}[18:24:31, 18:39:30]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.7435	N/A
$\overline{p_{dyn}^{(ptop)}}[18:39:30, 19:13:53]$	Mean dynamic power (PowerTOP) during the service instance's operation	15.5243	N/A

Table XXIV: scaleLevel40: processes in descending order of mean power use, up to 90th percentile of total

Description	PW Estimate (mW)
tick_sched_timer	4038.354
[3] net_rx(softirq)	1449.757
hrtimer_wakeup	935.0115
[PID 3882] containerd --config /var/run/docker/containerd/containerd.toml	251.189
[PID 3794] /usr/bin/dockerd	103.5928
[PID 18] [rcu_preempt]	68.97527
[PID 3881] containerd --config /var/run/docker/containerd/containerd.toml	37.20248
[PID 3892] containerd --config /var/run/docker/containerd/containerd.toml	29.134
toggle_allocation_gate	28.56159
[PID 3897] containerd --config /var/run/docker/containerd/containerd.toml	24.13863
[7] sched(softirq)	23.6729
[PID 8262] ffmpeg -re -ss 1711 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	20.19329
[PID 3893] containerd --config /var/run/docker/containerd/containerd.toml	19.71883
[PID 9089] ffmpeg -re -ss 247 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	19.58201
[PID 11433] ffmpeg -re -ss 415 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	19.01707
[PID 11453] ffmpeg -re -ss 2276 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.89512
[PID 11485] ffmpeg -re -ss 151 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.88354
[PID 8576] ffmpeg -re -ss 1644 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.86067
[PID 11942] ffmpeg -re -ss 2373 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.59756
[PID 11634] ffmpeg -re -ss 2897 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.54939
[PID 11469] ffmpeg -re -ss 427 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.54207
[PID 3884] containerd --config /var/run/docker/containerd/containerd.toml	18.42221
[PID 11798] ffmpeg -re -ss 632 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.28902
[PID 11514] ffmpeg -re -ss 1661 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.2753
[PID 11531] ffmpeg -re -ss 1299 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.24451
[PID 11938] ffmpeg -re -ss 271 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.24238
[PID 11562] ffmpeg -re -ss 220 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.2375
[PID 11628] ffmpeg -re -ss 109 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.21738
[PID 11604] ffmpeg -re -ss 2363 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.19177
[PID 11694] ffmpeg -re -ss 1300 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.15854
[PID 11700] ffmpeg -re -ss 580 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.15457

Description	PW Estimate (mW)
[PID 11810] ffmpeg -re -ss 2017 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.0625
[PID 11834] ffmpeg -re -ss 2577 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.06006
[PID 11670] ffmpeg -re -ss 712 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.05732
[PID 11500] ffmpeg -re -ss 2878 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.05091
[PID 11882] ffmpeg -re -ss 2350 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.04512
[PID 11754] ffmpeg -re -ss 2185 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.0378
[PID 11724] ffmpeg -re -ss 522 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.03659
[PID 3891] containerd --config /var/run/docker/containerd/containerd.toml	18.01934
[PID 11556] ffmpeg -re -ss 2037 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.01768
[PID 11748] ffmpeg -re -ss 493 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.01372
[PID 11822] ffmpeg -re -ss 1415 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.0061
[PID 11524] ffmpeg -re -ss 2590 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	18.00335
[PID 11804] ffmpeg -re -ss 2668 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.99939
[PID 11718] ffmpeg -re -ss 1344 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.97652
[PID 11712] ffmpeg -re -ss 1130 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.96677
[PID 11664] ffmpeg -re -ss 966 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.96006
[PID 11761] ffmpeg -re -ss 864 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.95122
[PID 11870] ffmpeg -re -ss 1546 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.95091
[PID 11888] ffmpeg -re -ss 2028 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.94878
[PID 11876] ffmpeg -re -ss 1359 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.94543
[PID 11846] ffmpeg -re -ss 1062 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.92713
[PID 11592] ffmpeg -re -ss 765 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.92256
[PID 11646] ffmpeg -re -ss 321 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.92104
[PID 11706] ffmpeg -re -ss 1270 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.91799
[PID 11840] ffmpeg -re -ss 217 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.91433
[PID 11598] ffmpeg -re -ss 683 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.9125
[PID 11622] ffmpeg -re -ss 2799 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.90671
[PID 11568] ffmpeg -re -ss 559 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.89604
[PID 11894] ffmpeg -re -ss 2495 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.87134
[PID 11688] ffmpeg -re -ss 2492 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.86341
[PID 11682] ffmpeg -re -ss 2593 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.85976
[PID 11586] ffmpeg -re -ss 1251 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.84146
[PID 11544] ffmpeg -re -ss 2668 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.83963
[PID 11574] ffmpeg -re -ss 1297 -i /videos/chosen.mp4 -t 1800 -c copy -f mpegts pipe:1	17.83049

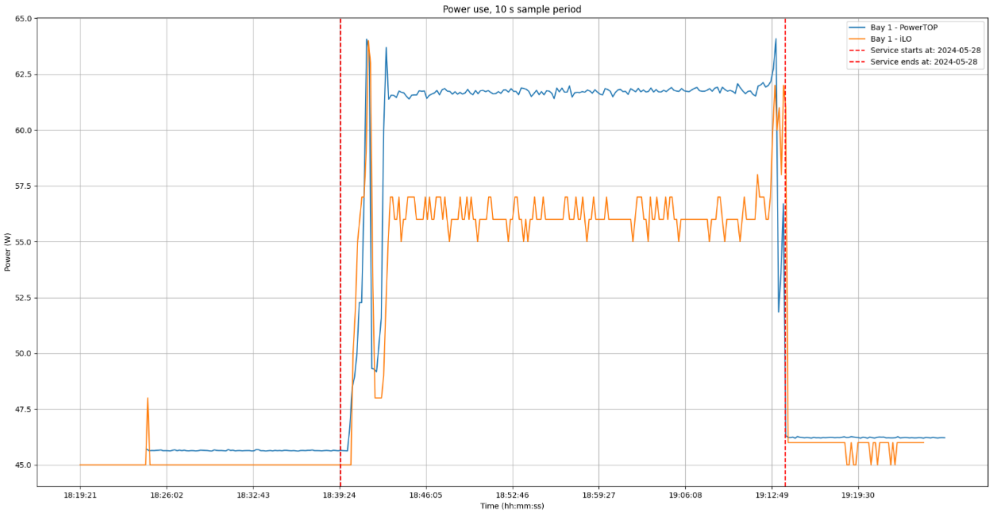


Fig. 40: scaleLevel 80. Video server’s power use during containerized service operation.

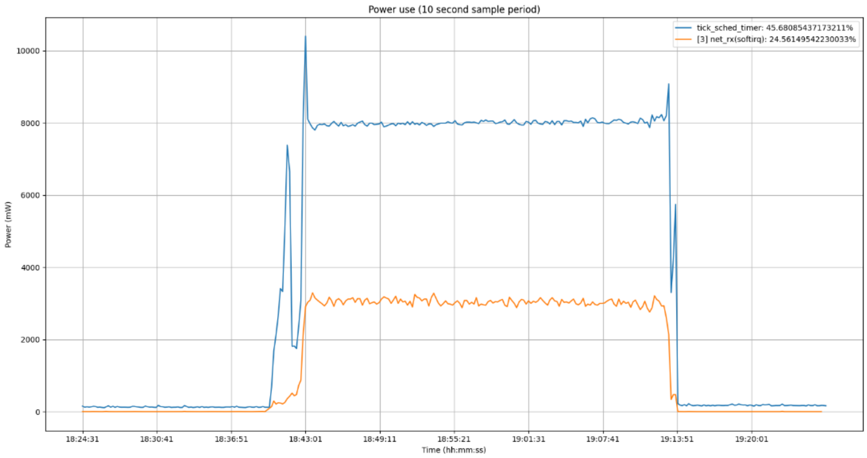


Fig. 41: scaleLevel 80. Process power use during automated containerized service operation, 50th percentile

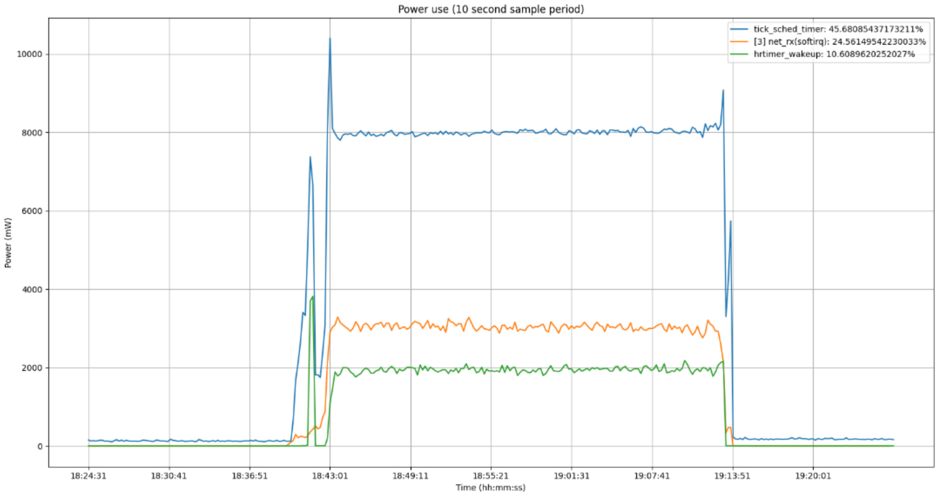
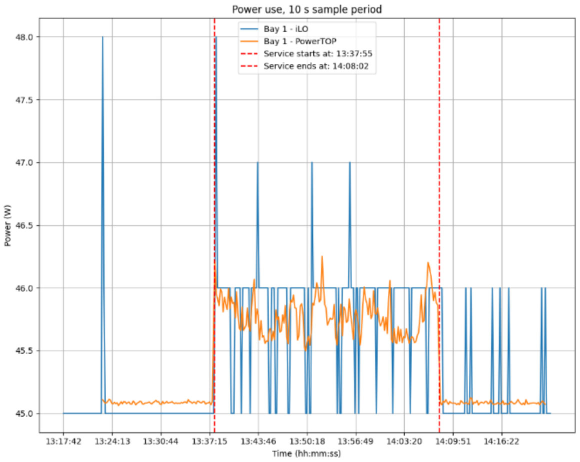
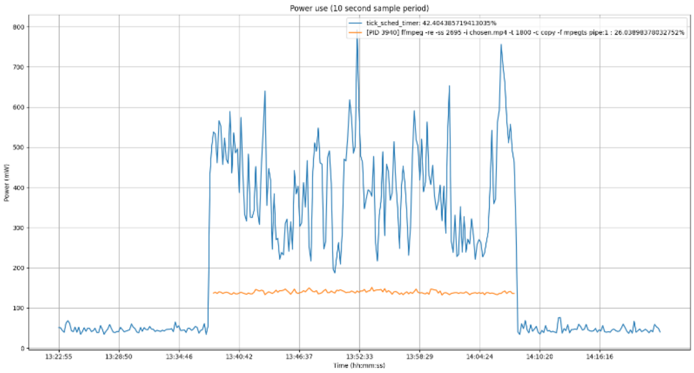
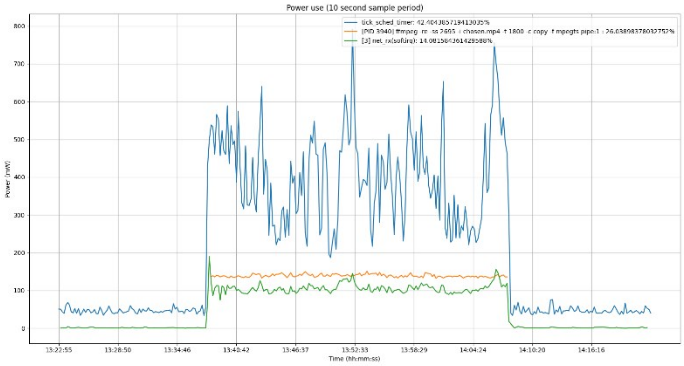


Fig. 42: scaleLevel 80. Process power use during automated containerized service operation, 80th percentile

6.6.3 Automated service management: Orchestration of native streaming

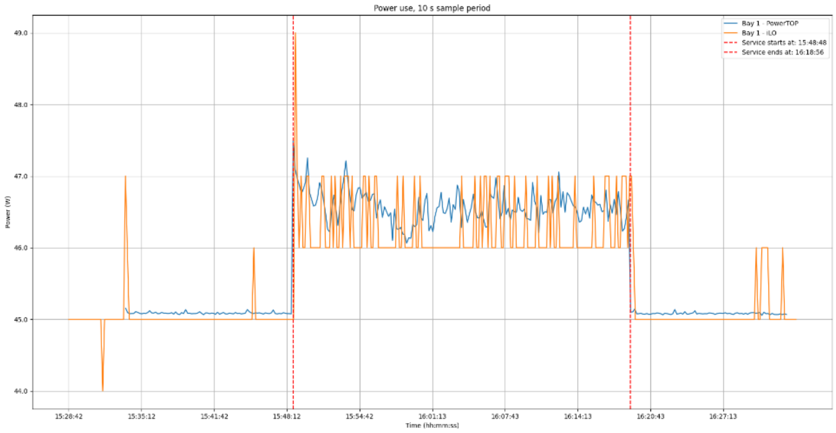
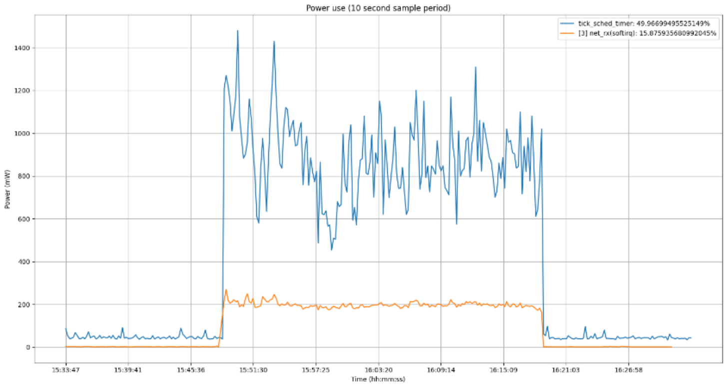
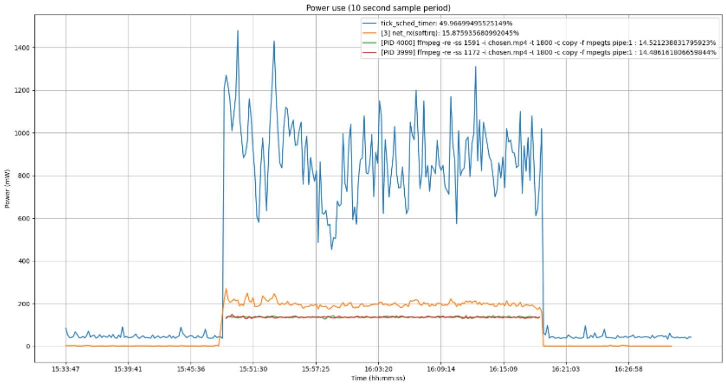
Single instance

Table XXV

<table>
<tr><th>Power type</th><th>Description</th><th>Average (W)</th><th>Range (W)</th></tr>
<tr><td>$\overline{p^{(iLO)}}[13:17:42,13:37:55]$</td><td>Before starting a service instance</td><td>45.54</td><td>[45.04,46.04]</td></tr>
<tr><td>$\overline{p^{(iLO)}}[13:37:55,14:08:02]$</td><td>During the service instance’s operation</td><td>46.38</td><td>[45.88,46.88]</td></tr>
<tr><td>$\overline{p_{dyn}^{(ptop)}}[13:22:55,13:37:55]$</td><td>Mean dynamic power (PowerTOP) before the service instance’s operation</td><td>0.2080</td><td>N/A</td></tr>
<tr><td>$\overline{p_{dyn}^{(ptop)}}[13:37:55,14:08:02]$</td><td>Mean dynamic power (PowerTOP) during the service instance’s operation</td><td>0.8675</td><td>N/A</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">Fig. 43: Native streaming, 1 stream. Total power, 50th percentile and 80th percentile (top to bottom)</td></tr>
</table>

Two instances

Table XXVI

<table>
<tr><th>Power type</th><th>Description</th><th>Average (W)</th><th>Range (W)</th></tr>
<tr><td>$\overline{p^{(iLO)}}[15{:}28{:}42,15{:}48{:}48]$</td><td>Before starting a service instance</td><td>45.525</td><td>[45.025,46.025]</td></tr>
<tr><td>$\overline{p^{(iLO)}}[15{:}48{:}48,16{:}18{:}56]$</td><td>During the service instance’s operation</td><td>46.79</td><td>[46.29,47.29]</td></tr>
<tr><td>$\overline{p_{dyn}^{(ptop)}}[15{:}33{:}47,15{:}48{:}48]$</td><td>Mean dynamic power (PowerTOP) before the service instance’s operation</td><td>0.2390</td><td>N/A</td></tr>
<tr><td>$\overline{p_{dyn}^{(ptop)}}[15{:}48{:}48,16{:}18{:}56]$</td><td>Mean dynamic power (PowerTOP) during the service instance’s operation</td><td>1.6653</td><td>N/A</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">Fig. 44: Native streaming, 2 streams. Total power, 50th percentile and 90th percentile (top to bottom)</td></tr>
</table>

Five instances

Table XXVII

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[17{:}49{:}44,18{:}09{:}58]$	Before starting a service instance	45.5	[45.0,46.0]
$\overline{p^{(iLO)}}[18{:}09{:}58,18{:}40{:}10]$	During the service instance's operation	**47.71**	**[47.21,48.21]**
$\overline{p_{dyn}^{(ptop)}}[17{:}54{:}57,18{:}09{:}58]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.2546	N/A
$\overline{p_{dyn}^{(ptop)}}[18{:}09{:}58,18{:}40{:}10]$	Mean dynamic power (PowerTOP) during the service instance's operation	3.7906	N/A

Fig. 45: Native streaming, 5 streams. Total power, 50th percentile and 90th percentile (top to bottom)

Ten instances

Table XXVIII

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[19:21:06, 19:41:12]$	Before starting a service instance	45.54	[45.04,46.04]
$\overline{p^{(iLO)}}[19:41:12, 20:11:32]$	During the service instance's operation	**49.33**	**[48.83,49.83]**
$\overline{p_{dyn}^{(ptop)}}[19:26:11, 19:41:12]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.2466	N/A
$\overline{p_{dyn}^{(ptop)}}[19:41:12, 20:11:32]$	Mean dynamic power (PowerTOP) during the service instance's operation	7.4315	N/A

Fig. 46: Native streaming, 10 streams. Total power, 50th percentile and **95th** percentile (top to bottom)

Twenty instances

Table XXIX

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[20{:}52{:}00, 21{:}12{:}09]$	Before starting a service instance	45.52	[45.02,46.02]
$\overline{p^{(iLO)}}[21{:}12{:}09, 21{:}42{:}48]$	During the service instance's operation	**51.27**	**[50.77,51,77]**
$\overline{p_{dyn}^{(ptop)}}[20{:}57{:}09, 21{:}12{:}09]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.1819	N/A
$\overline{p_{dyn}^{(ptop)}}[21{:}12{:}09, 21{:}42{:}48]$	Mean dynamic power (PowerTOP) during the service instance's operation	14.3948	N/A

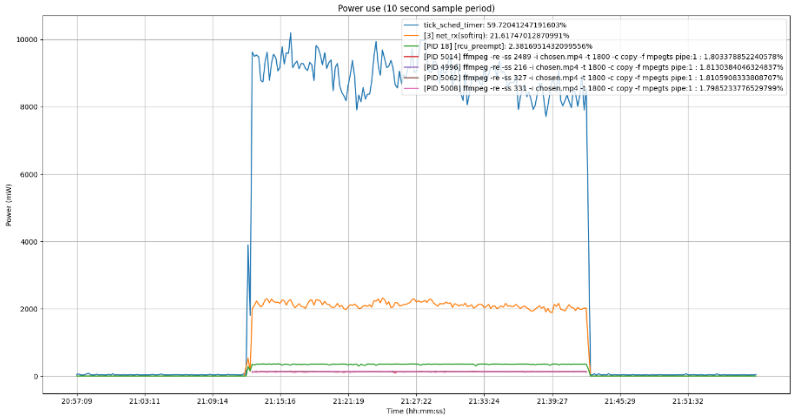
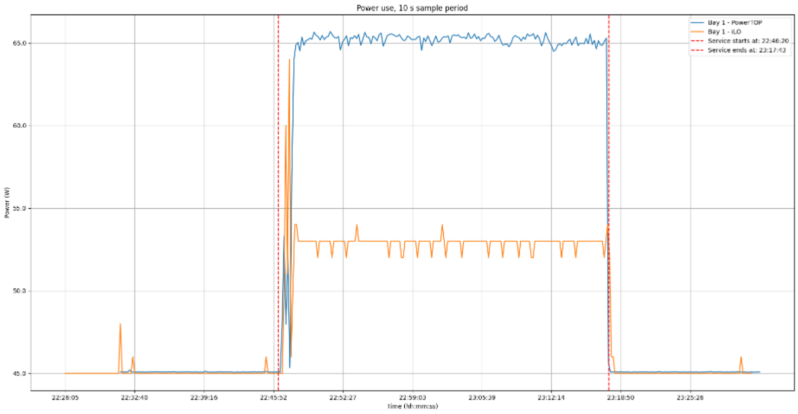
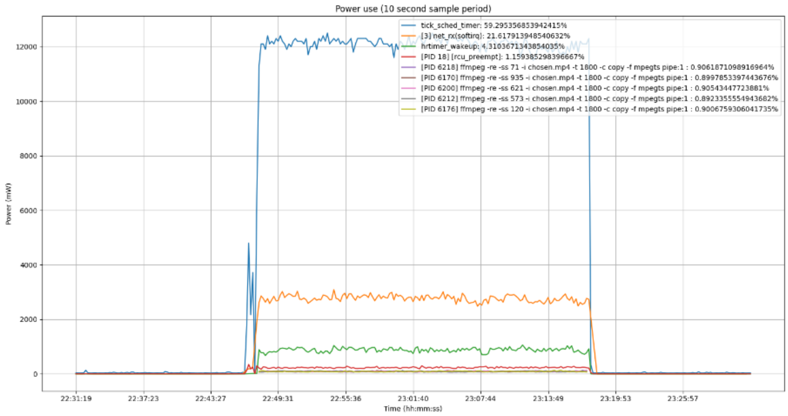


Fig. 47: Native streaming, 20 streams. Total power, **75th** percentile and **90th** percentile (top to bottom)

Forty instances

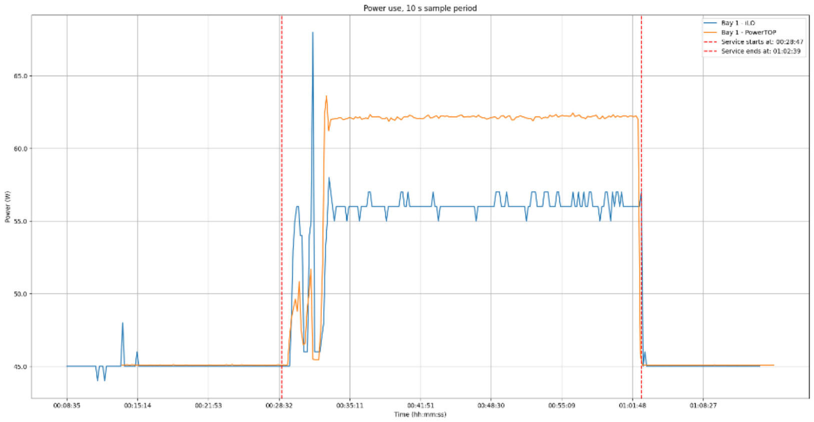
Table XXX

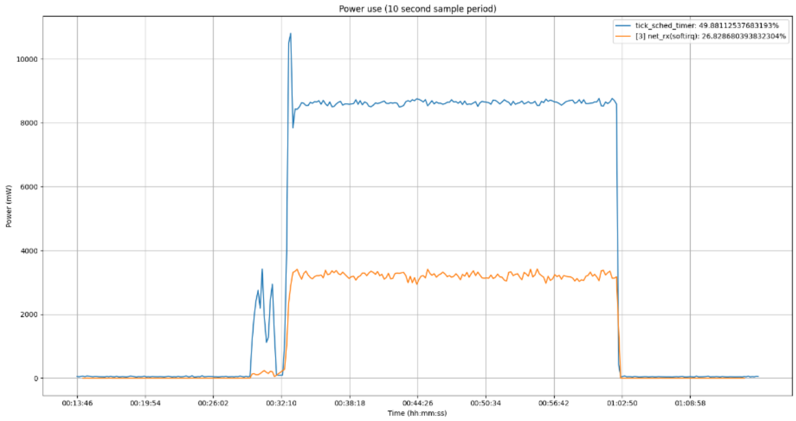
<table>
<tr><th>Power type</th><th>Description</th><th>Average (W)</th><th>Range (W)</th></tr>
<tr><td>$\overline{p^{(iLO)}}[22:26:05,22:46:20]$</td><td>Before starting a service instance</td><td>45.54</td><td>[45.04,46.04]</td></tr>
<tr><td>$\overline{p^{(iLO)}}[22:46:20,23:17:43]$</td><td>During the service instance’s operation</td><td>**53.27**</td><td>**[52.77,53.77]**</td></tr>
<tr><td>$\overline{p_{dyn}^{(ptop)}}[22:31:19,22:46:20]$</td><td>Mean dynamic power (PowerTOP) before the service instance’s operation</td><td>0.1780</td><td>N/A</td></tr>
<tr><td>$\overline{p_{dyn}^{(ptop)}}[22:46:20,23:17:43]$</td><td>Mean dynamic power (PowerTOP) during the service instance’s operation</td><td>19.2853</td><td>N/A</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">
</td></tr>
<tr><td colspan="4">Fig. 48: Native streaming, 40 streams. Total power, 75th percentile and 90th percentile (top to bottom)</td></tr>
</table>

Eighty instances

Table XXXI

Power type	Description	Average (W)	Range (W)
$\overline{p^{(iLO)}}[00{:}08{:}35, 00{:}28{:}47]$	Before starting a service instance	45.52	[45.02,46.02]
$\overline{p^{(iLO)}}[00{:}28{:}47, 01{:}02{:}39]$	During the service instance's operation	**55.81**	**[55.31,56.31]**
$\overline{p_{dyn}^{(ptop)}}[00{:}13{:}46, 00{:}28{:}47]$	Mean dynamic power (PowerTOP) before the service instance's operation	0.1874	N/A
$\overline{p_{dyn}^{(ptop)}}[00{:}28{:}47, 01{:}02{:}39]$	Mean dynamic power (PowerTOP) during the service instance's operation	15.1443	N/A





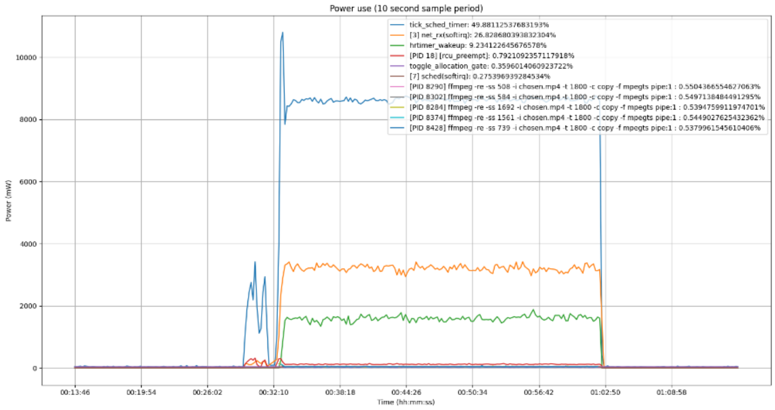


Fig. 49: Native streaming, 80 streams. Total power, **75th** percentile and **90th** percentile (top to bottom)

7. Discussion

7.1 Preliminary: reconciliation of results as a basis for further experimentation

Table XXXII summarizes the results obtained[77] for the cases of manually managed and automated containerized operation respectively. The difference between mean dynamic power figures is significant and warrants investigation.

Table XXXII

Power type	**Description**	**Average (W)**	**Range (W)**
MANUALLY MANAGED OPERATION			
$\overline{p^{(iLO)}}[13{:}34{:}23,14{:}04{:}14]$	Before starting a service instance	45.01	[44.51,45.51]
$\overline{p^{(iLO)}}[14{:}04{:}14,15{:}36{:}16]$	During the service instance's operation	**46.06**	**[45.56,46.56]**
$\overline{p^{(iLO)}}[15{:}36{:}16,15{:}47{:}50]$	After the service instance ended	45.63	[45.13,46.13]
$\overline{p_{dyn}^{(ptop)}}[14{:}04{:}14,15{:}36{:}16]$	Mean dynamic power (PowerTOP) during the service instance's operation	2.1295	N/A
AUTOMATED OPERATION			
$\overline{p^{(iLO)}}[12{:}04{:}58,12{:}25{:}08]$	Before starting a service instance	45.05	[44.55,45.55]
$\overline{p^{(iLO)}}[12{:}25{:}08,13{:}57{:}19]$	During the service instance's operation	**45.78**	**[45.28,46.28]**
$\overline{p^{(iLO)}}[13{:}57{:}19,14{:}10{:}09]$	After the service instance ended	45.69	[45.19,46.19]
$\overline{p_{dyn}^{(ptop)}}[12{:}25{:}08,13{:}57{:}19]$	Mean dynamic power (PowerTOP) during the service instance's operation	1.3845	N/A

The results have been investigated by comparing the ordered process power means for both cases. Table XXXIII shows a consistently higher power estimate for the manually-managed case, across all processes. The cause is unclear, but the mutually conciliatory results for the automated cases suggest a good modus operandi (i.e., that which employed Ansible for automation) that can be adopted (see section 6.6.2) to exercise a platform over a range of loads.

7.2 Growth of power use with number of streams

Various characterizations of power use are considered and plotted in Fig. 50 – Fig. 52. Notation is shown below; the (n) symbol indicates dependence of power used on number of streaming containers.

1. total power during operations, $P_{ops}^{iLO}(n)$;
2. dynamic power during operations, $P_{ops}^{dyn}(n)$;
3. differential total power between operations and quiescence, $P_{ops}^{iLO}(n) - P_{q}^{iLO}$, and
4. differential dynamic power between operations and quiescence, $P_{ops}^{dyn}(n) - P_{q}^{dyn}$;

---

[77] The experiment was repeated, for verification, and the same quantitative difference was observed.

Fig. 53 shows the difference in power use between the two streaming platforms. The non-monotonic behaviour is due to the error introduced by the rounding of iLO instrumentation.

7.3 Evaluation of work carried out.

The objective set out in section 3 was to quantify the overhead incurred by operating the video service containerized, instead of as an application running directly on the host operating system (native operation). An access network of the Active Ethernet type was constructed and a video cache deployed in an access node to stream videos to the access node's service area. An implementation model describing the access network was included.

The results obtained have shown that the overhead is negligible and that the benefit of running the video source in a container comes at little cost. The possibility of consolidating video streaming containers can be pursued with confidence.

No discernable cause for concern was found in the power measurement instrumentation embedded in the HPE Gen9 platform. Documentation on interfacing with the Integrated Lights-Out (iLO) server management was readily available. For detail beyond typical interest, HPE readily divulged information on this tool when contacted for help, including, for example, the method used to round the power measurement into an integer [381].

On the other hand, PowerTOP's accuracy poses a problem. The various graphs of power against time have shown that it captures changes well, but gravely overestimates them. In the light of these errors, works that have investigated containerization's overhead with the use of this tool (e.g., [374]) must be reviewed for the implications of inaccuracies introduced by the tool.

Baselines have been obtained for both the video server and the virtual switch. In particular, P_{b1}^{video} has been found useful in providing an offset for power obtained through tools that measure dynamic power. This segues well into an observation that merits particular attention. Even with 80 concurrent streams, the static power has dwarfed the dynamic power. The importance of this observation pertains to the importance of the benefit of containerization as an enabler of consolidation of physical hosts. It can readily be stated that the overhead incurred in providing the **service framework** of containerization, poses no obstacle to exploration of exploitation of this benefit.

Table XXXIII

Manual management		Automated management	
Description	Power (mW)	Description	Power (mW)
tick_sched_timer	453	tick_sched_timer	340
[PID 3849] containerd --config /var/run/docker/containerd/containerd.toml	355	[PID 3850] containerd --config /var/run/docker/containerd/containerd.toml	214
[3] net_rx(softirq)	323	[PID 4325] ffmpeg -re -i /videos/chosen.mp4 -c copy -f mpegts pipe:1	94
[PID 4199] /usr/bin/containerd-shim-runc-v2 -namespace moby -id 4618343bd39e3412ee6c5ee32fea672f1d0491bf23ecd7cd8b51ce2ee6f1488	111	[3] net_rx(softirq)	74
[PID 4232] ffmpeg -re -i ./chosen.mp4 -c:v copy -f mpegts tcp://10.0.0.1:7778?listen	107	[PID 3866] containerd --config /var/run/docker/containerd/containerd.toml	44
[PID 3859] containerd --config /var/run/docker/containerd/containerd.toml	69	toggle_allocation_gate	38
[PID 3850] containerd --config /var/run/docker/containerd/containerd.toml	69	[PID 18] [rcu_preempt]	36
[PID 3857] containerd --config /var/run/docker/containerd/containerd.toml	50	[PID 3856] containerd --config /var/run/docker/containerd/containerd.toml	34
[PID 3862] containerd --config /var/run/docker/containerd/containerd.toml	48	[PID 3863] containerd --config /var/run/docker/containerd/containerd.toml	32
[PID 17] [rcu_preempt]	48	[PID 3867] containerd --config /var/run/docker/containerd/containerd.toml	27
[PID 3861] containerd --config /var/run/docker/containerd/containerd.toml	46	[PID 3868] containerd --config /var/run/docker/containerd/containerd.toml	25
[PID 3858] containerd --config /var/run/docker/containerd/containerd.toml	44	[PID 3869] containerd --config /var/run/docker/containerd/containerd.toml	23
toggle_allocation_gate	38	[PID 3852] containerd --config /var/run/docker/containerd/containerd.toml	21
[PID 3865] containerd --config /var/run/docker/containerd/containerd.toml	35	fb_flashcursor	19
[PID 4230] ffmpeg -re -i ./chosen.mp4 -c:v copy -f mpegts tcp://10.0.0.1:7778?listen	28	[PID 4323] ffmpeg -re -i /videos/chosen.mp4 -c copy -f mpegts pipe:1	19

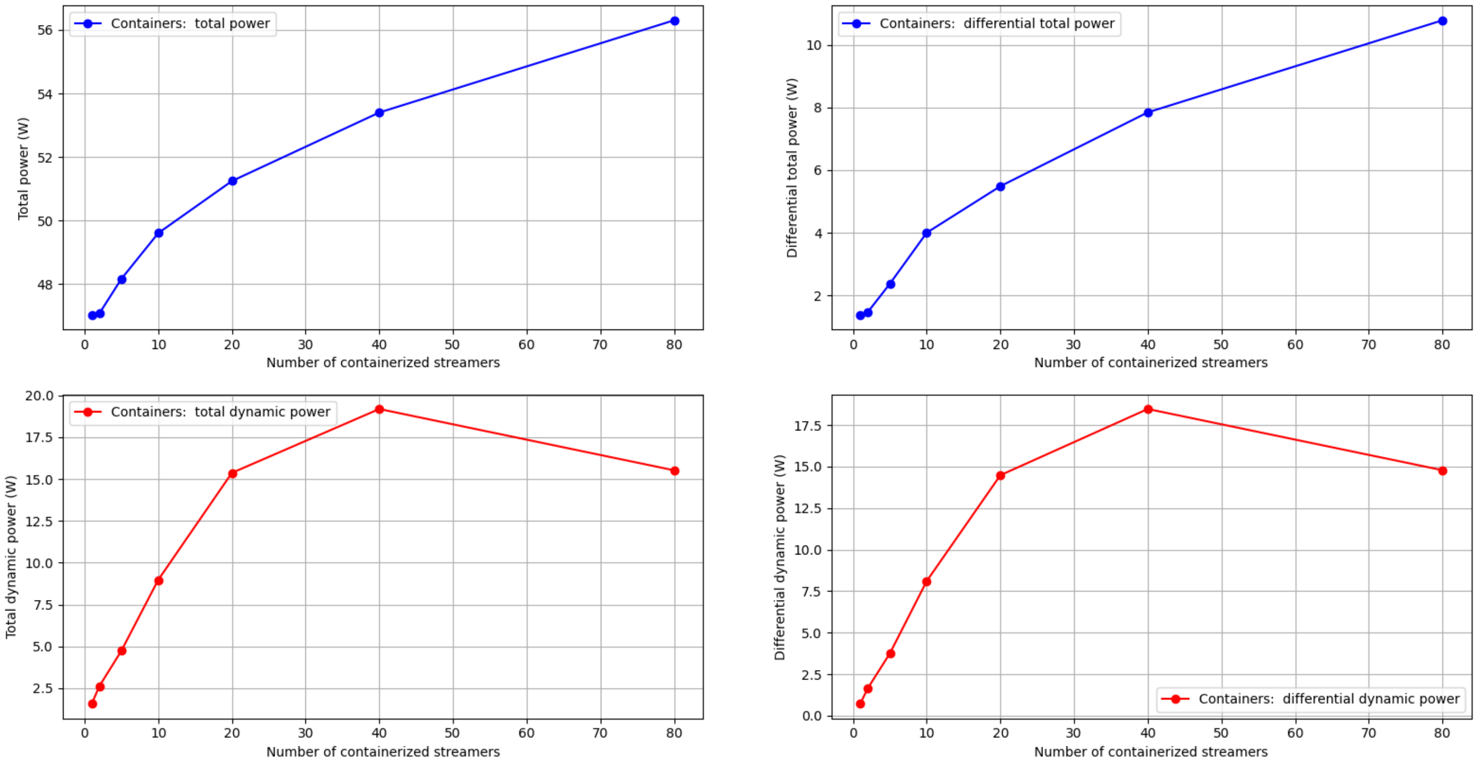


Fig. 50: Clockwise from top left: $P_{ops}^{iLO}(n)$, $P_{ops}^{iLO}(n) - P_q^{iLO}$, $P_{ops}^{dyn}(n)$ and $P_{ops}^{dyn}(n) - P_q^{dyn}$ for the case of containerized streaming

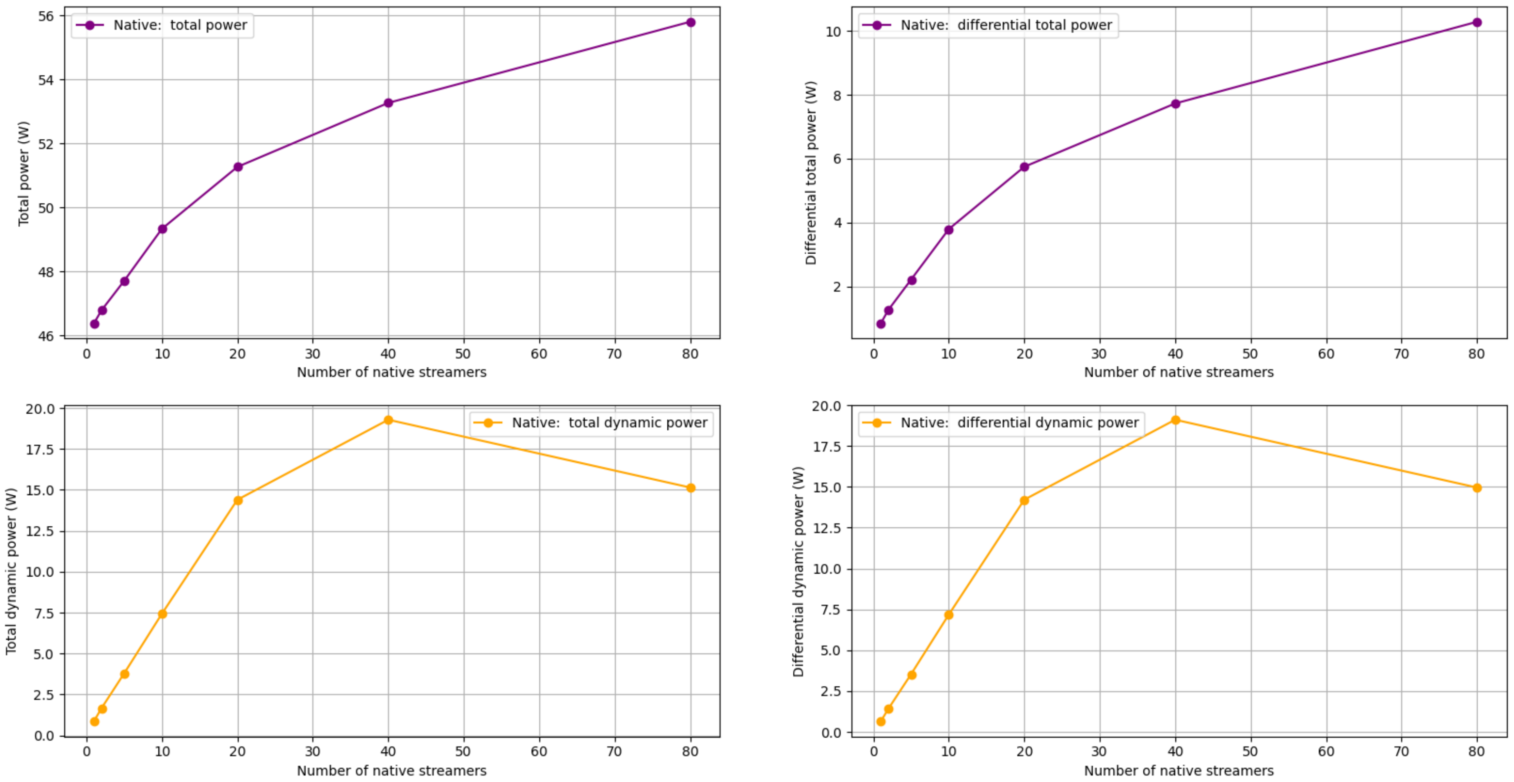


Fig. 51: Clockwise from top left: $P_{ops}^{iLO}(n)$, $P_{ops}^{iLO}(n) - P_{q}^{iLO}$, $P_{ops}^{dyn}(n)$ *and* $P_{ops}^{dyn}(n) - P_{q}^{dyn}$ for the case of native streaming

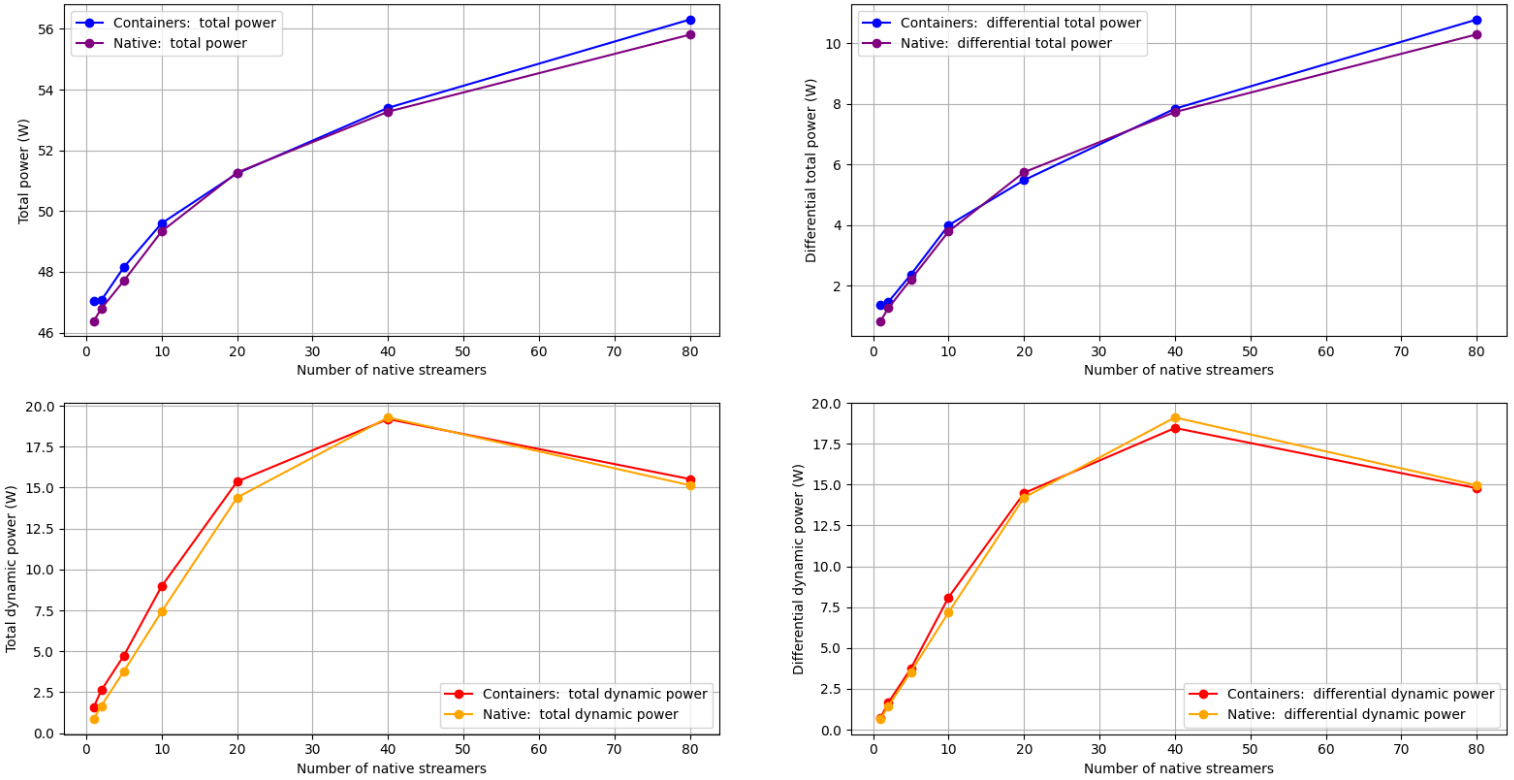


Fig. 52: Comparison: native vs containerized streaming. Clockwise from top left: $P_{ops}^{iLO}(n)$, $P_{ops}^{iLO}(n) - P_q^{iLO}$, $P_{ops}^{dyn}(n)$ and $P_{ops}^{dyn}(n) - P_q^{dyn}$

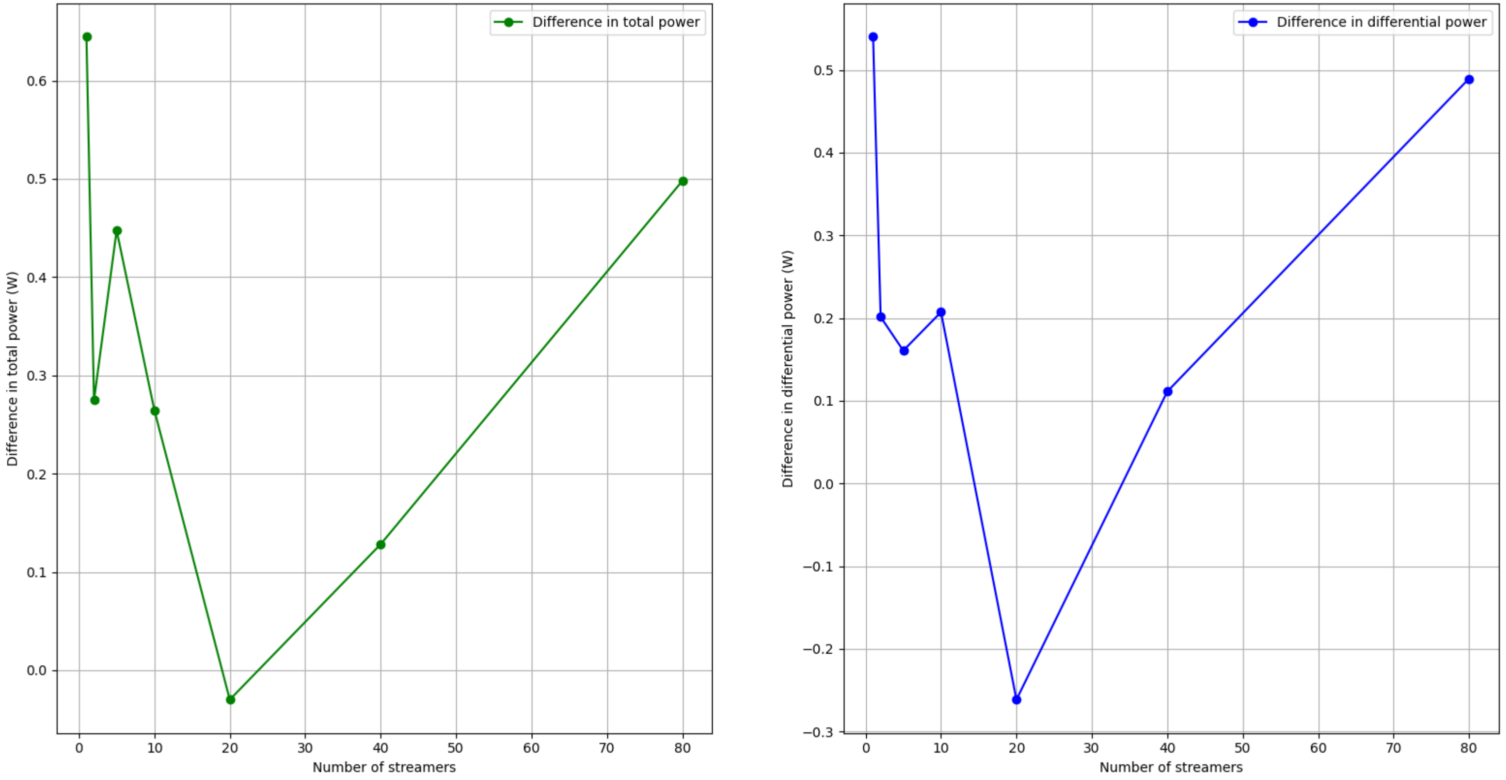


Fig. 53: Comparison: native vs containerized streaming. Difference in power use between the two streaming platforms

Appendix 8: Bitrate of video transmissions from containerized sources

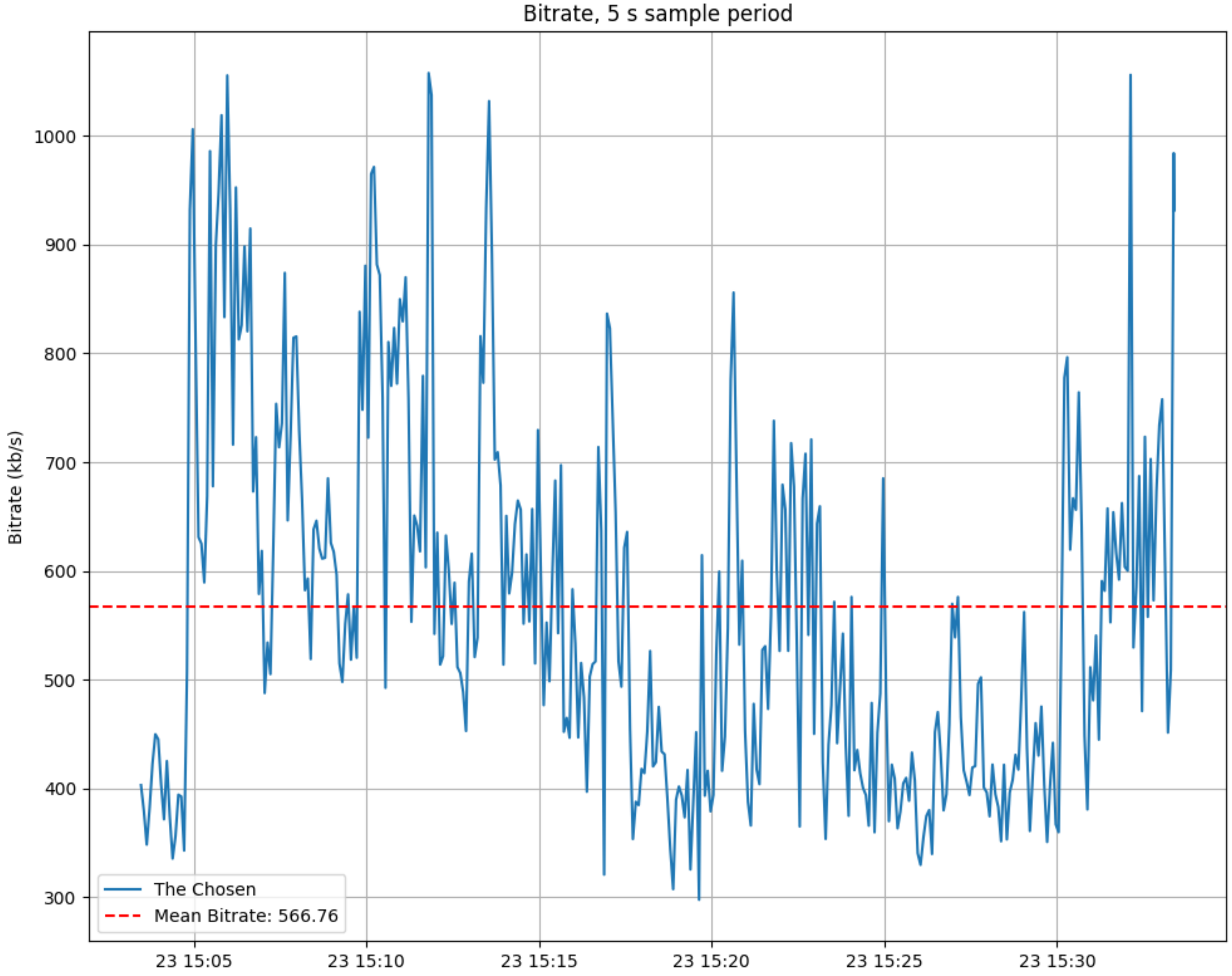


Fig. 54: Bitrate measured at receiver's end, while receiving one video stream. Dashed red line shows mean received bitrate.

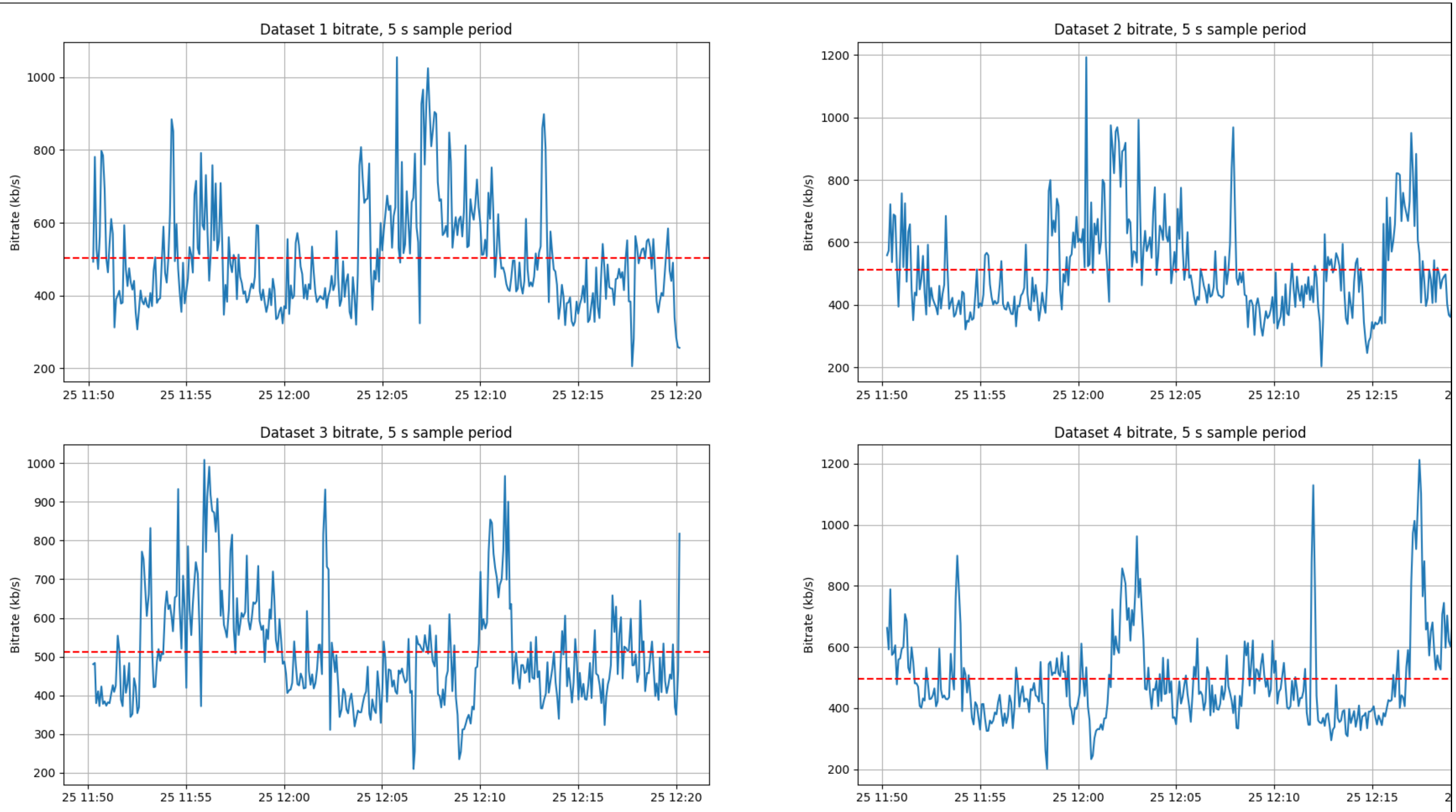


Fig. 55: Bitrate measured at receiver’s end, while receiving 5 video streams, streams 1 – 4. Dashed red line shows mean received bitrate.

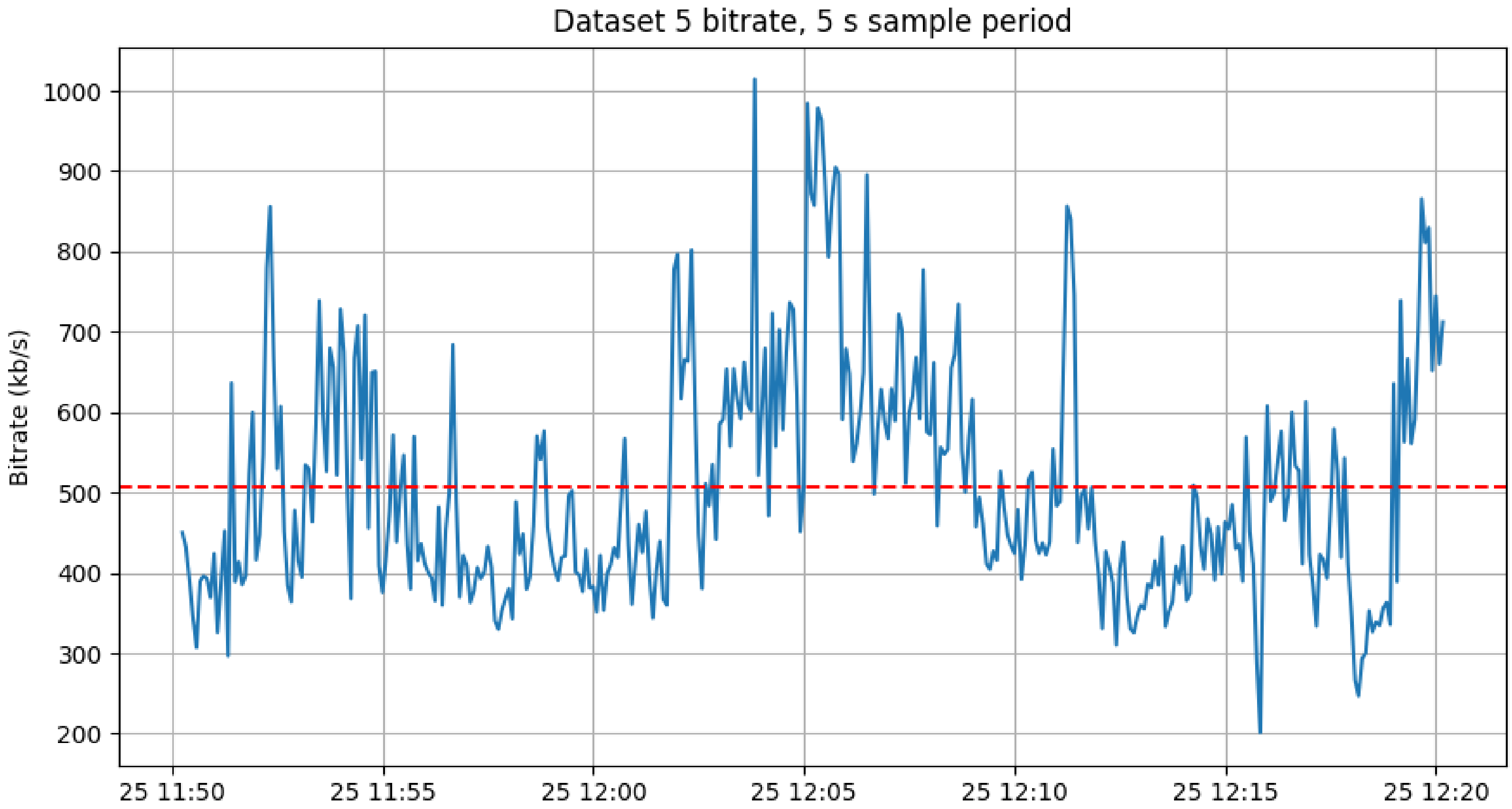


Fig. 56: Bitrate measured at receiver’s end, while receiving 5 video streams, stream 5. Dashed red line shows mean received bitrate.

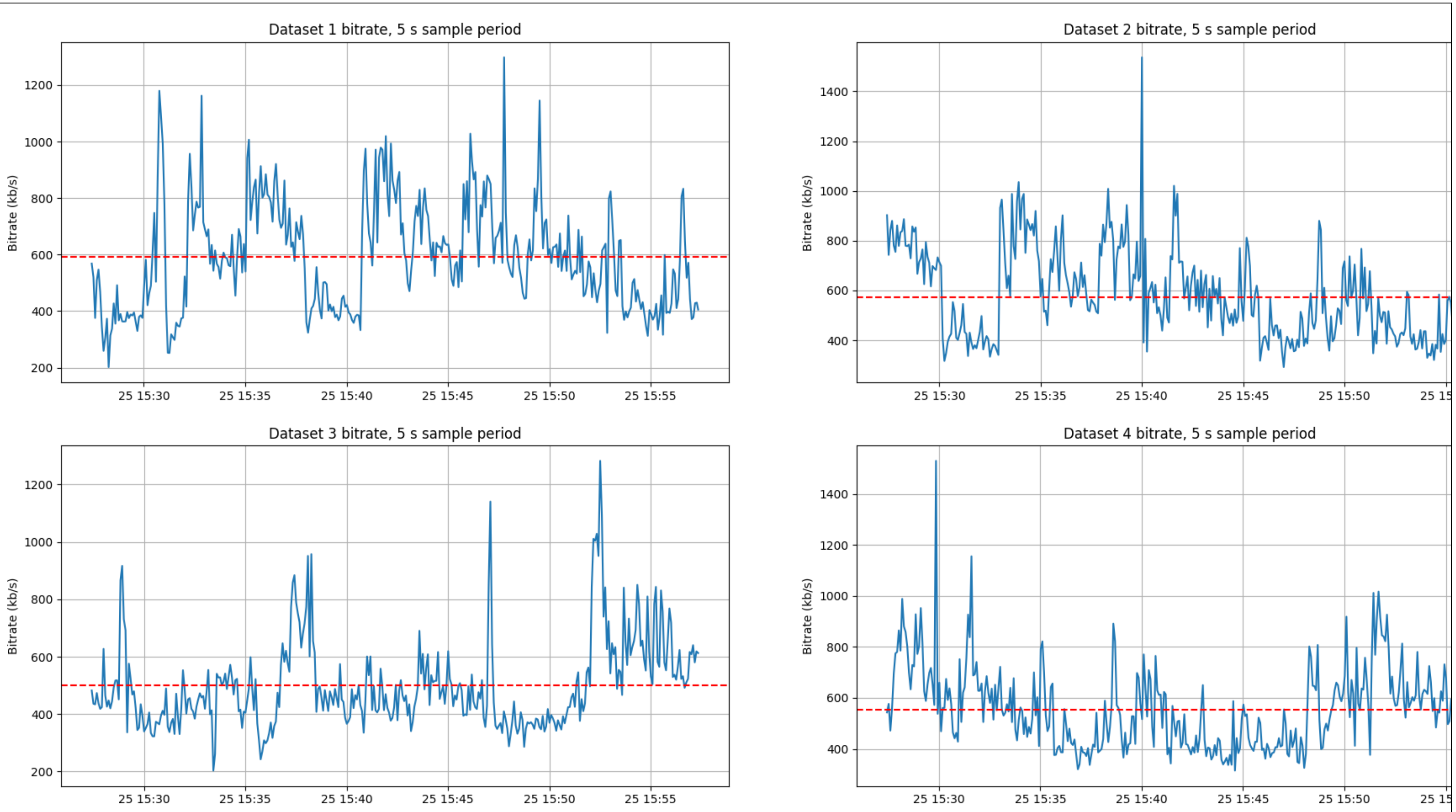


Fig. 57: Bitrate measured at receiver's end, while receiving 10 video streams, streams 1 – 4. Dashed red line shows mean received bitrate.

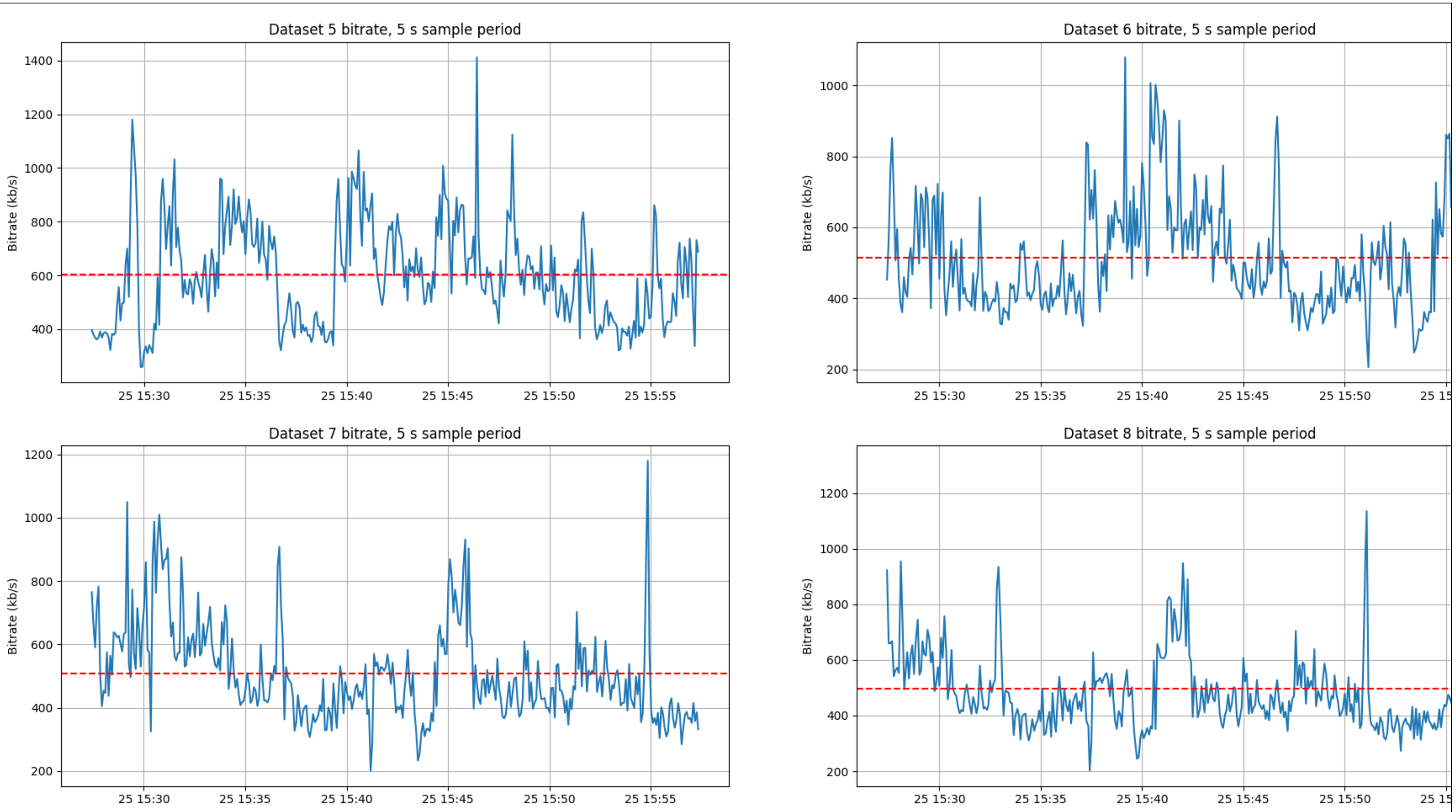


Fig. 58: Bitrate measured at receiver's end, while receiving 10 video streams, streams 5 – 8. Dashed red line shows mean received bitrate.

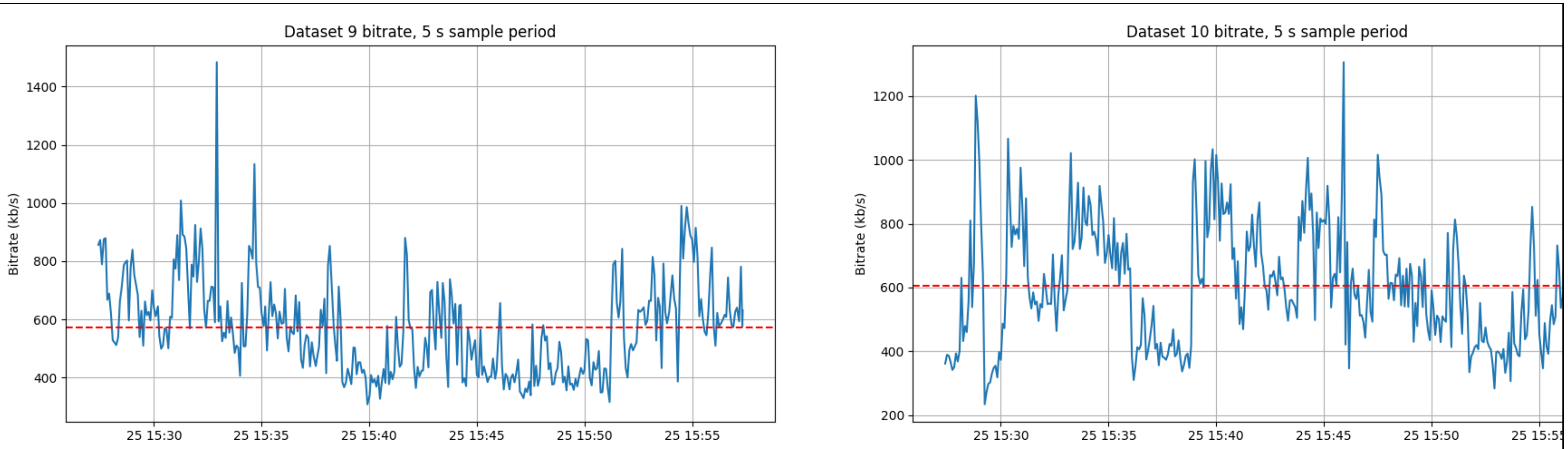


Fig. 59: Bitrate measured at receiver’s end, while receiving 10 video streams, streams 9 – 10. Dashed red line shows mean received bitrate.

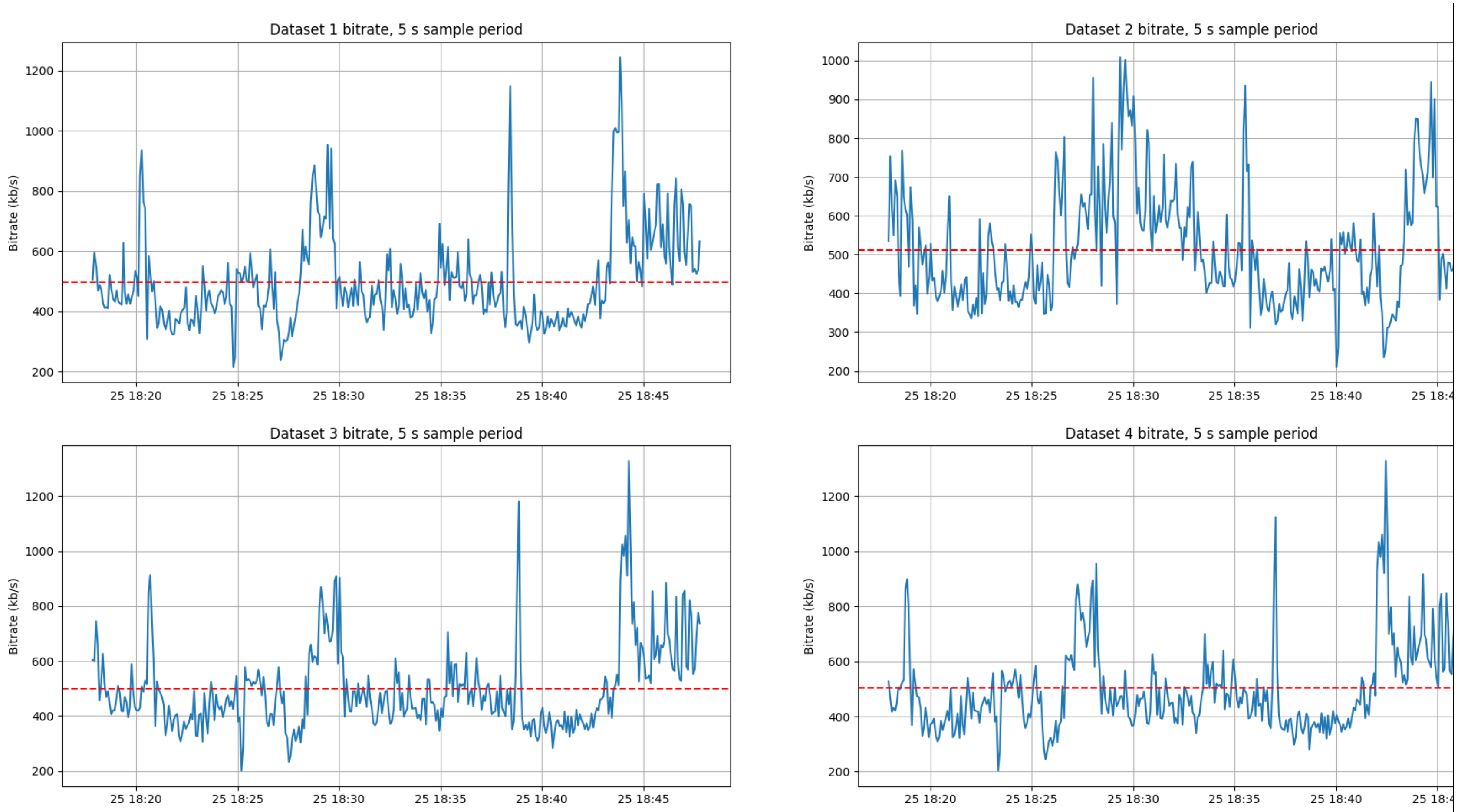


Fig. 60: Bitrate measured at receiver’s end, while receiving 20 video streams, streams 1 – 4. Dashed red line shows mean received bitrate.

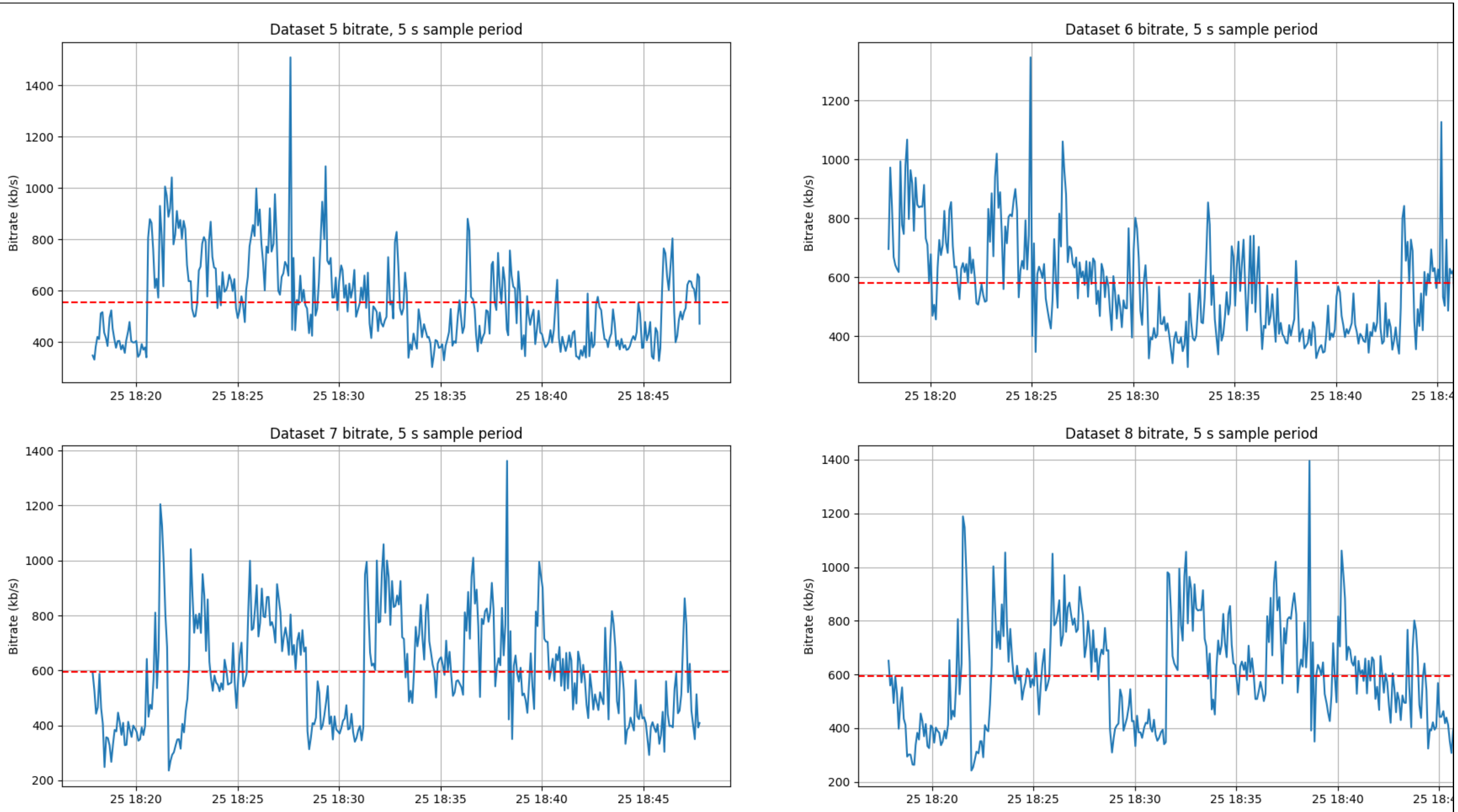


Fig. 61: Bitrate measured at receiver's end, while receiving 20 video streams, streams 5 – 8. Dashed red line shows mean received bitrate.

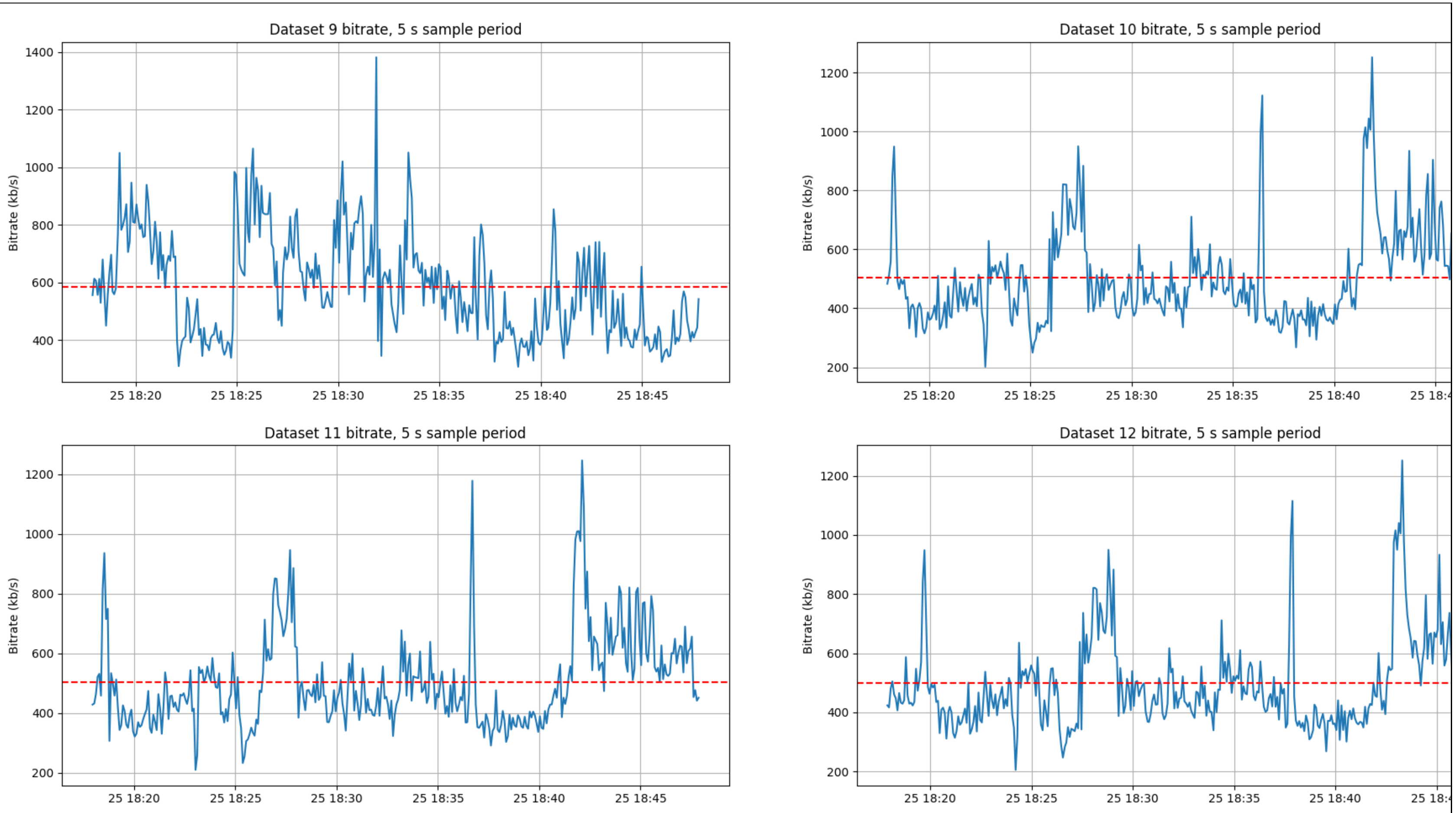


Fig. 62: Bitrate measured at receiver's end, while receiving 20 video streams, streams 9 – 12. Dashed red line shows mean received bitrate.

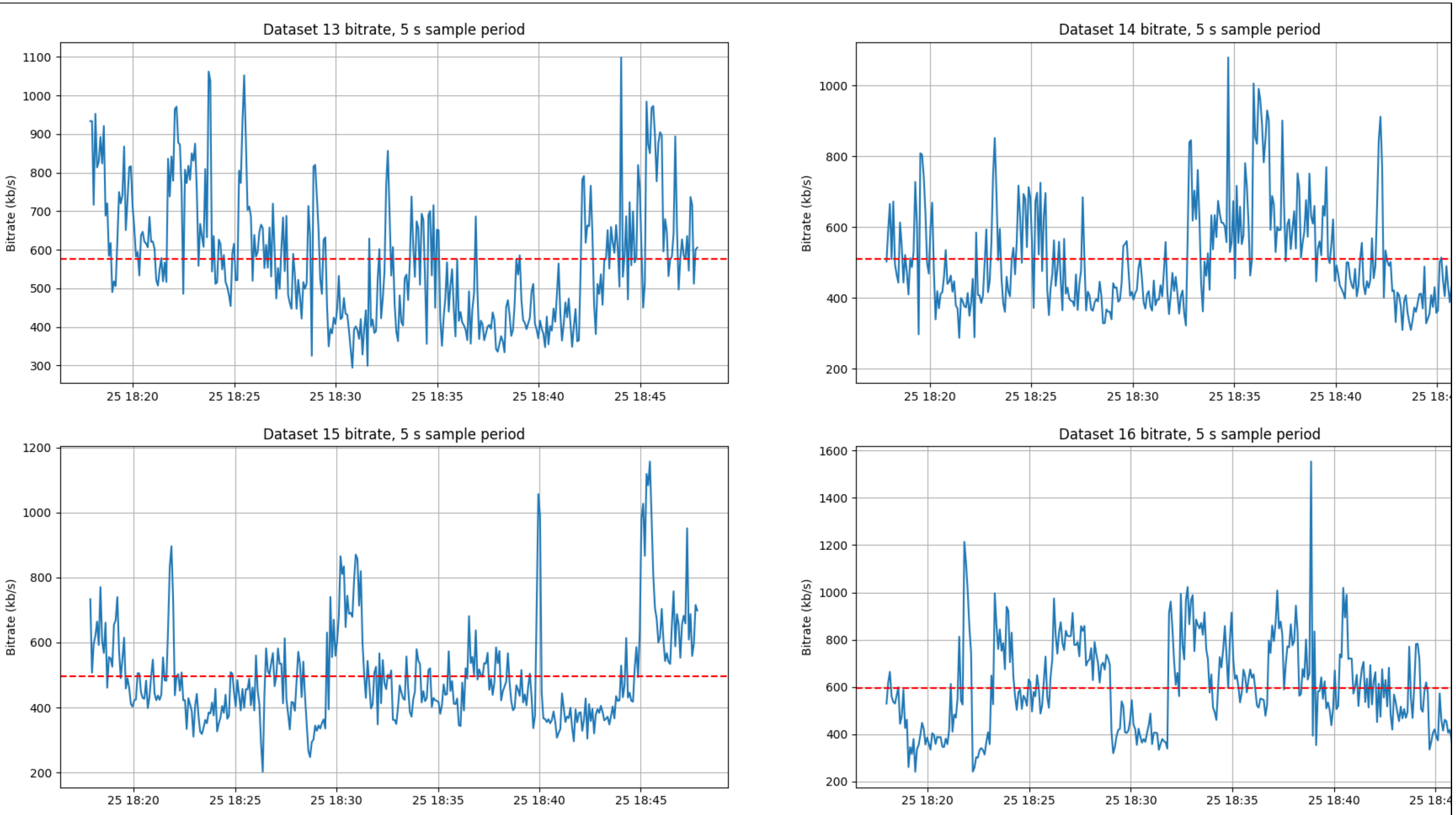


Fig. 63: Bitrate measured at receiver's end, while receiving 20 video streams, streams 13 – 16. Dashed red line shows mean received bitrate.

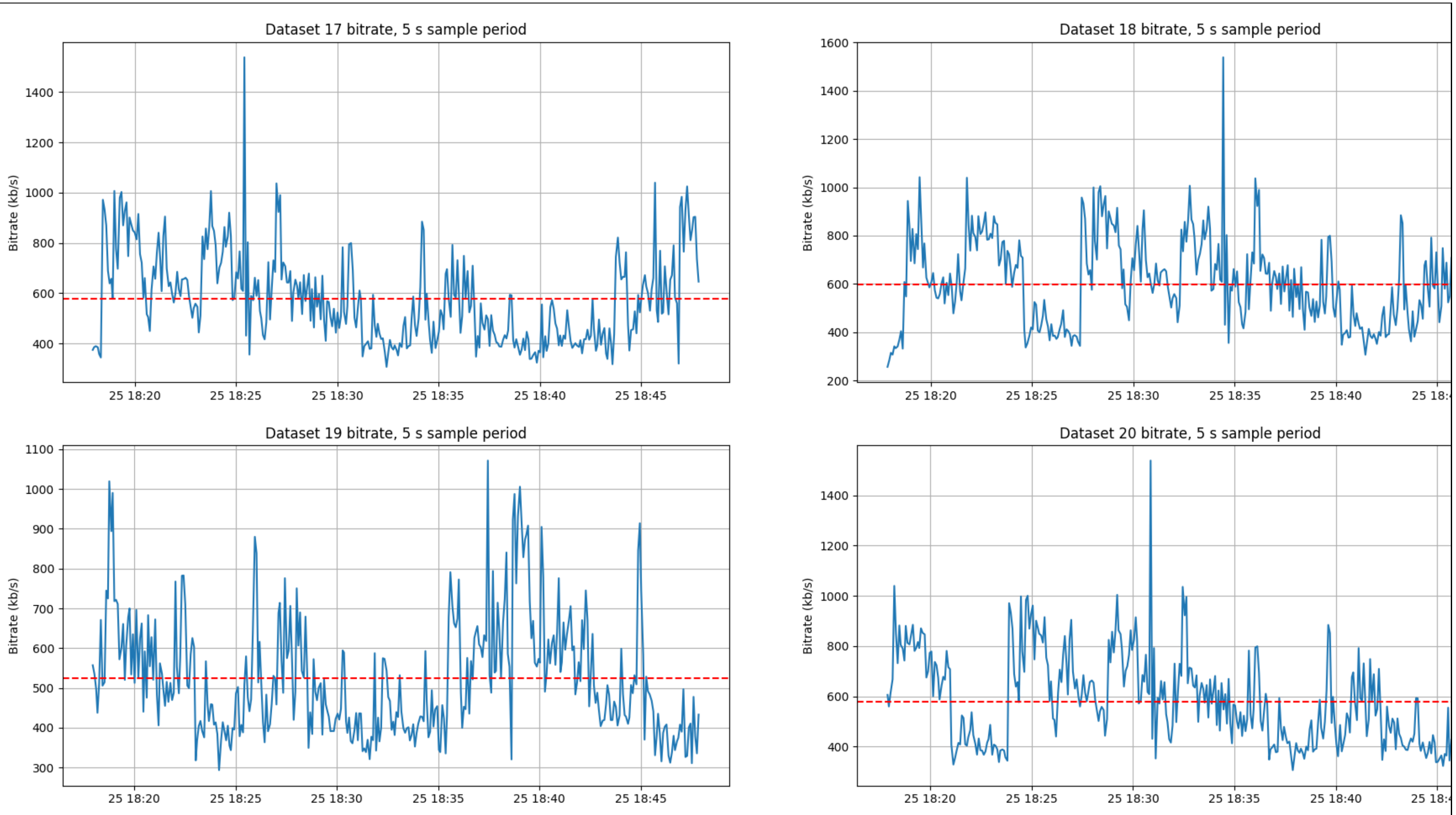


Fig. 64: Bitrate measured at receiver’s end, while receiving 20 video streams, streams 17 – 20. Dashed red line shows mean received bitrate.

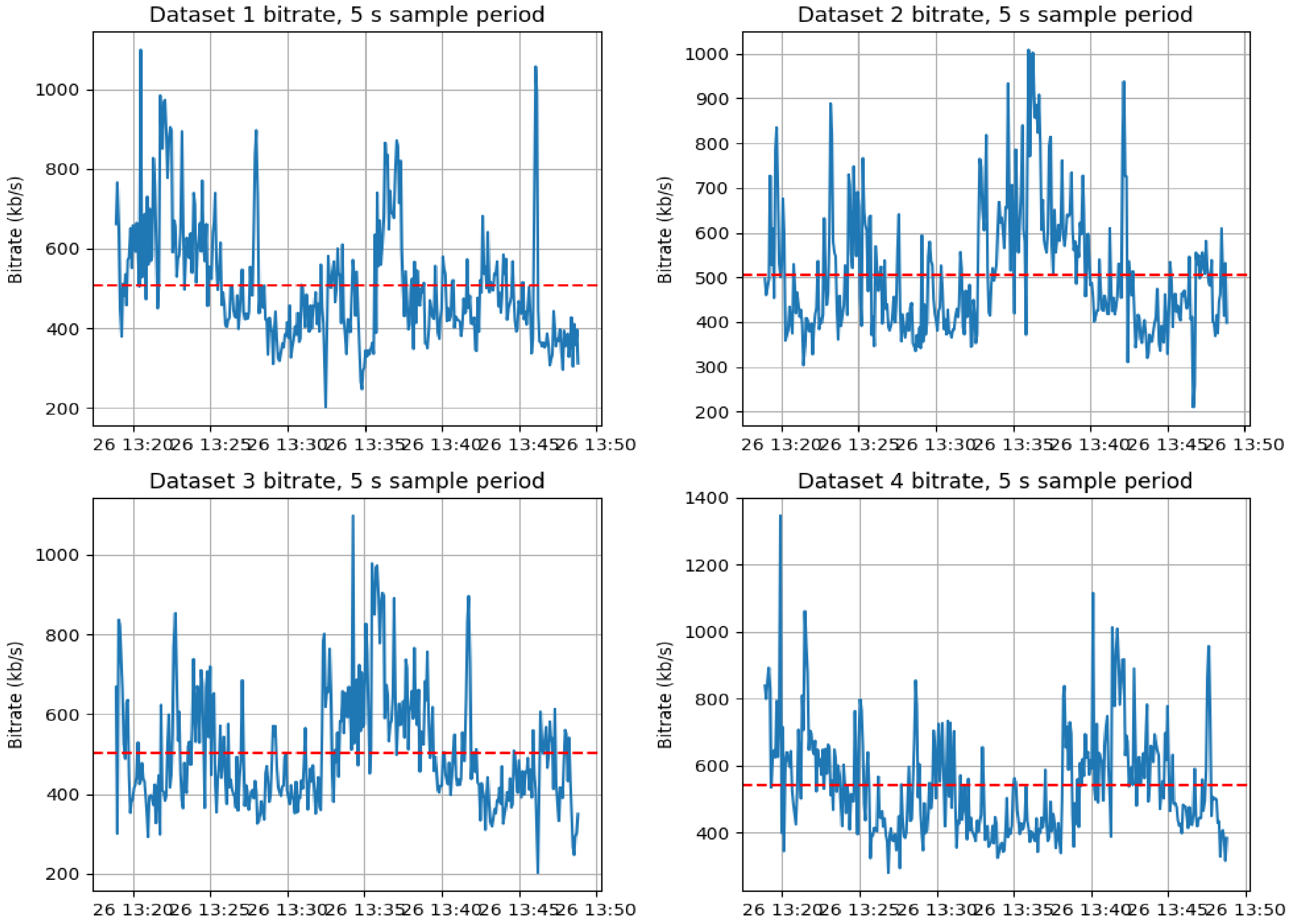


Fig. 65: Bitrate measured at receiver's end, while receiving 40 video streams, streams 1 – 4. Dashed red line shows mean received bitrate.

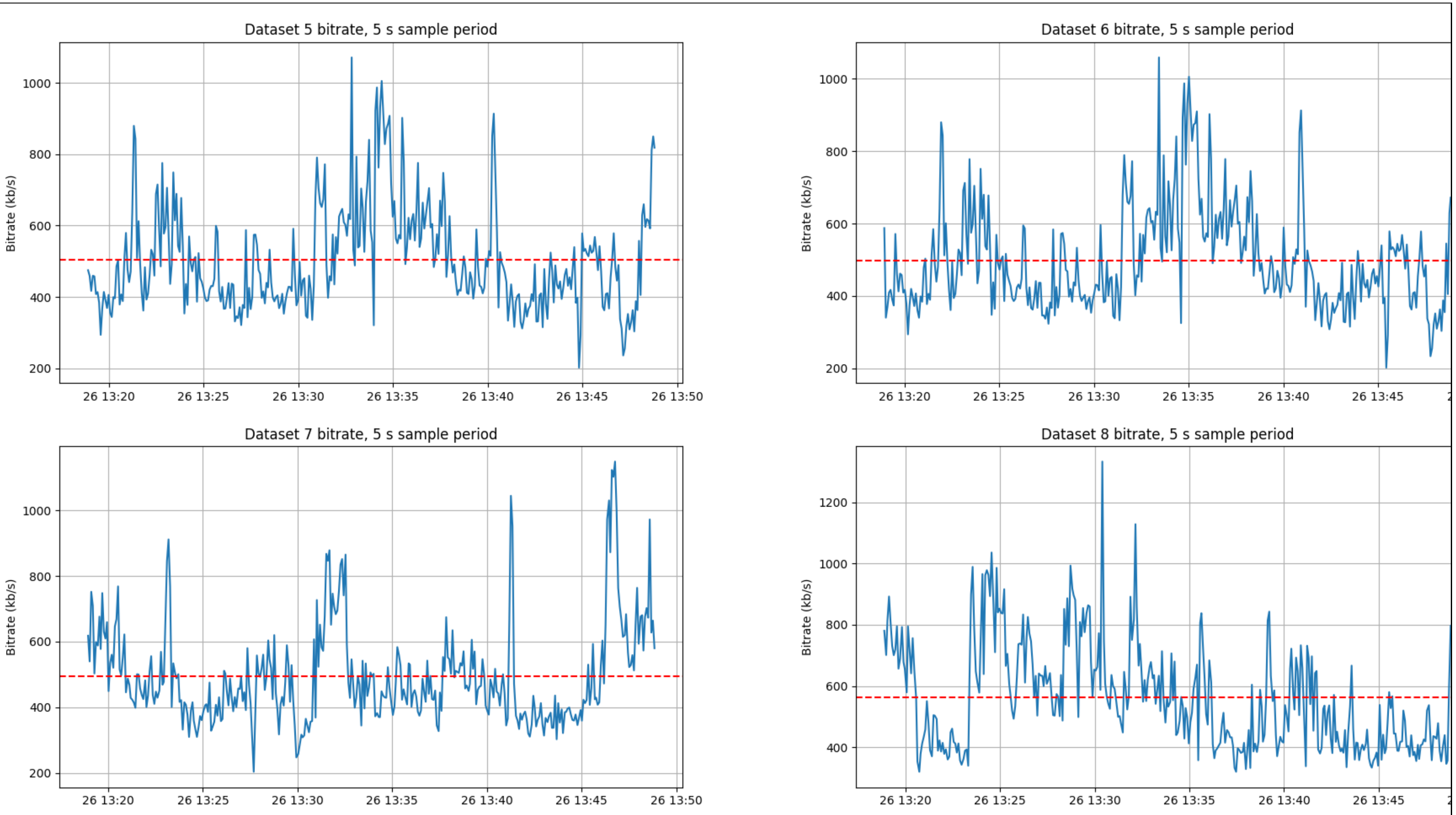


Fig. 66: Bitrate measured at receiver’s end, while receiving 40 video streams, streams 5 – 8. Dashed red line shows mean received bitrate.

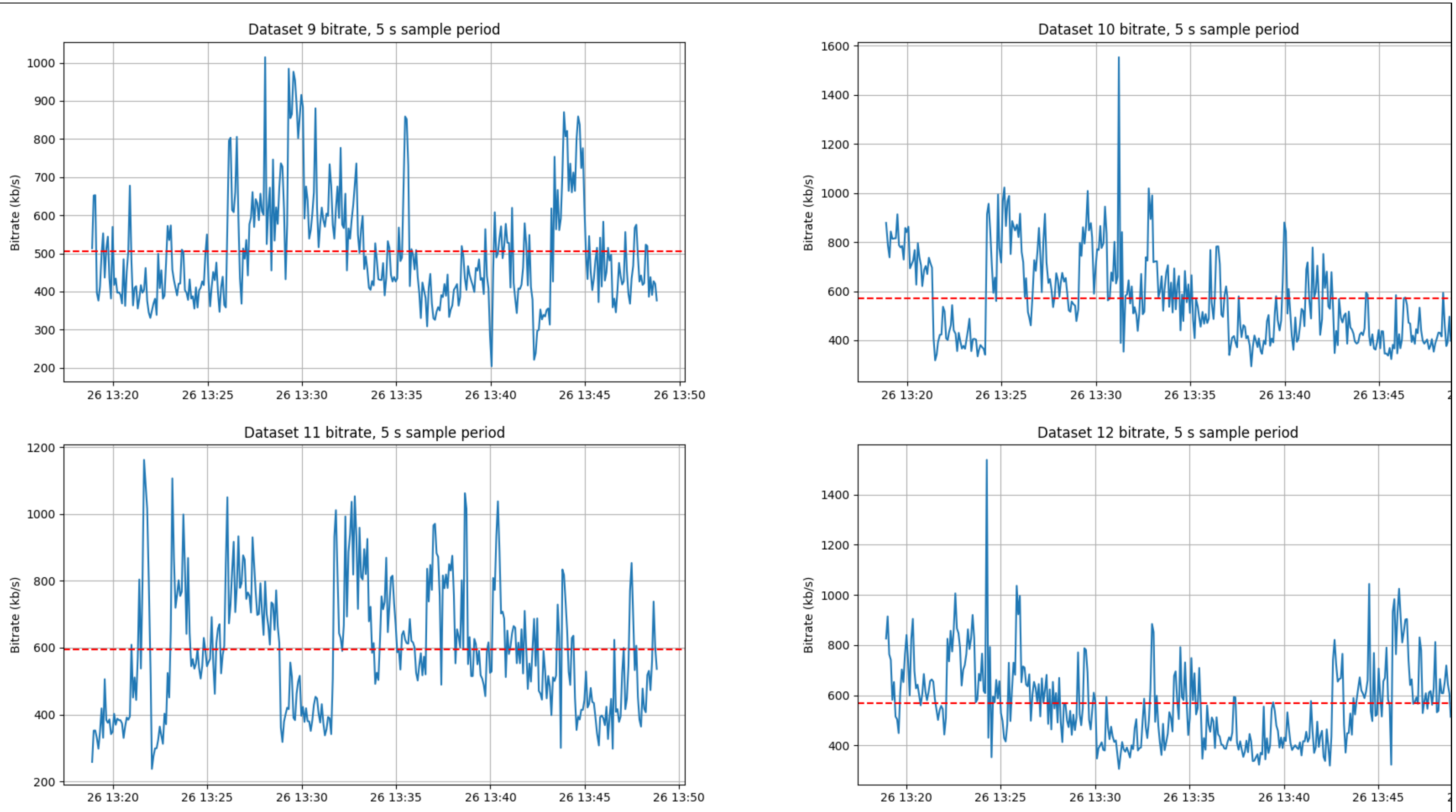


Fig. 67: Bitrate measured at receiver's end, while receiving 40 video streams, streams 9 – 12. Dashed red line shows mean received bitrate.

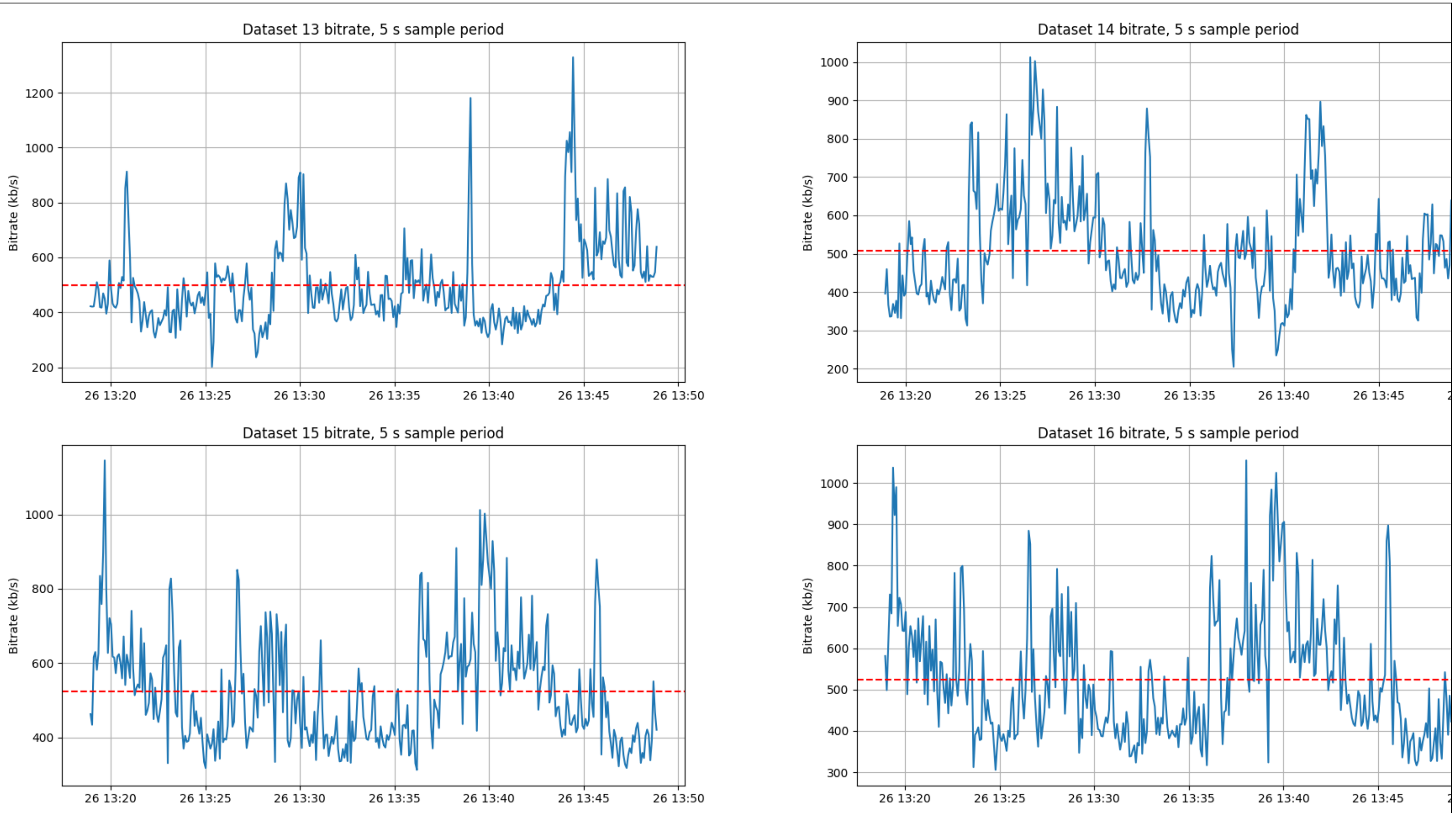
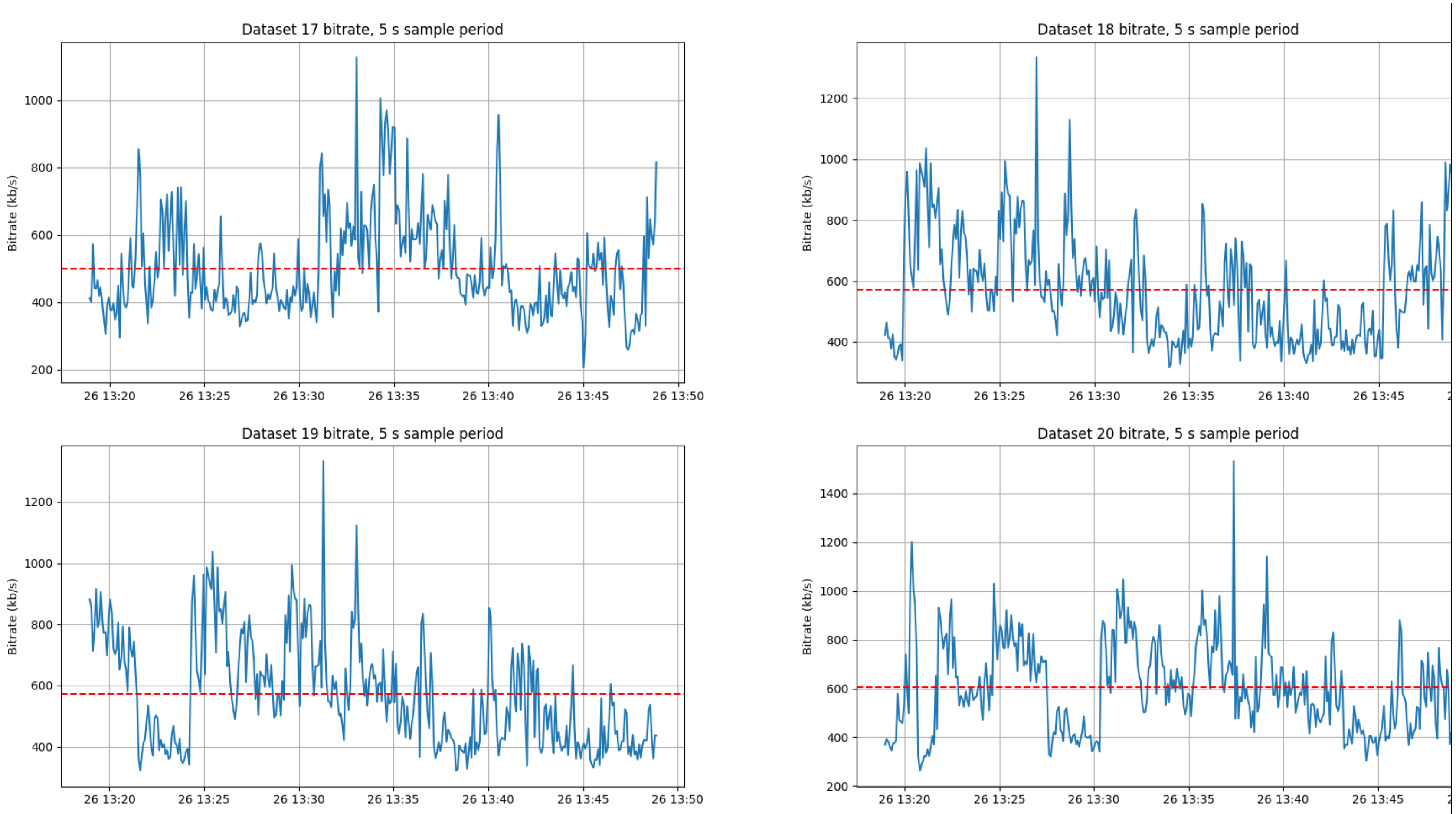


Fig. 68: Bitrate measured at receiver’s end, while receiving 40 video streams, streams 13 – 16. Dashed red line shows mean received bitrate.


Dataset 17 bitrate, 5 s sample period

Dataset 18 bitrate, 5 s sample period

Dataset 19 bitrate, 5 s sample period

Dataset 20 bitrate, 5 s sample period

Fig. 69: Bitrate measured at receiver’s end, while receiving 40 video streams, streams 17 – 20. Dashed red line shows mean received bitrate.

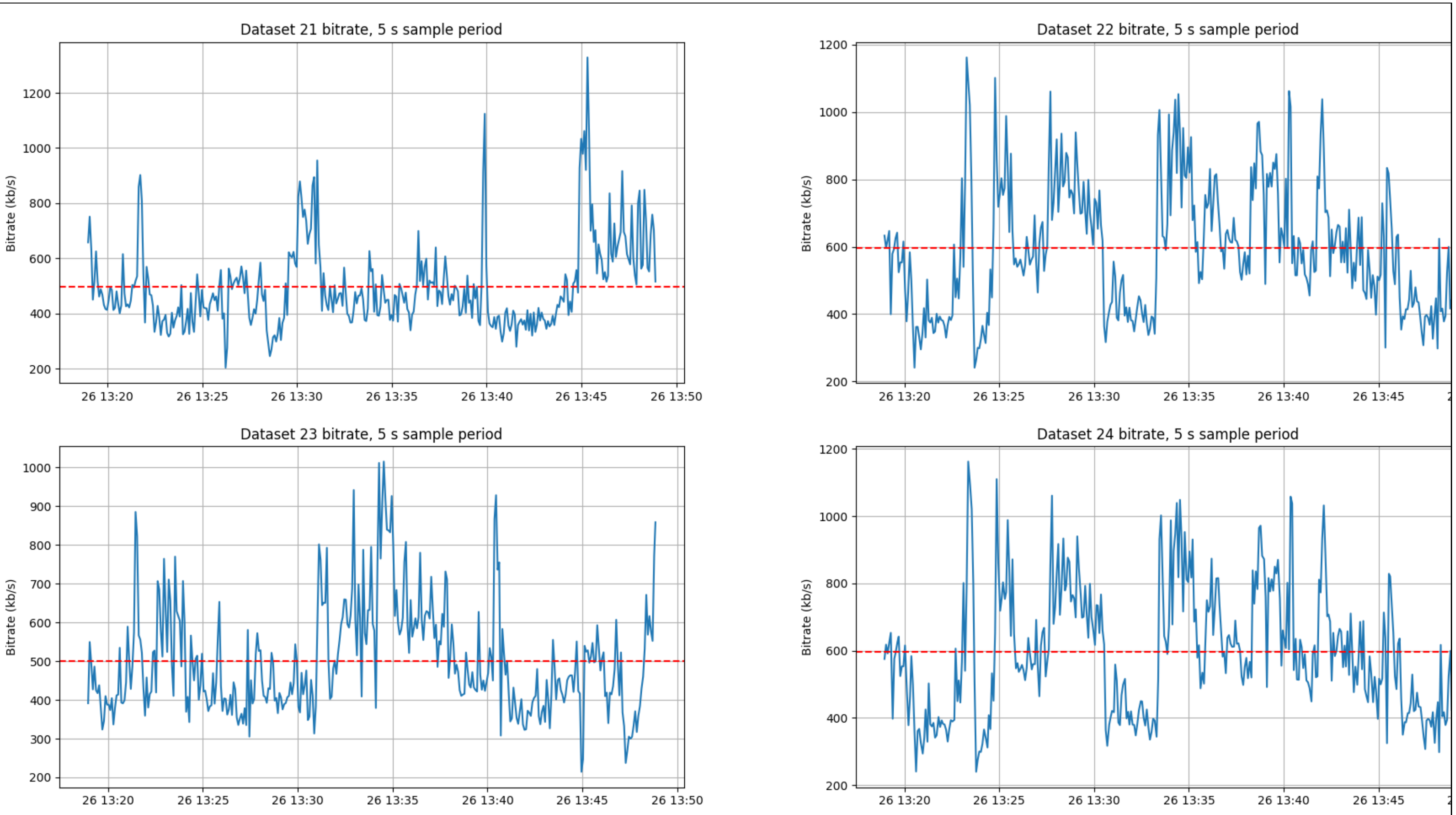


Fig. 70: Bitrate measured at receiver’s end, while receiving 40 video streams, streams 21 – 24. Dashed red line shows mean received bitrate.

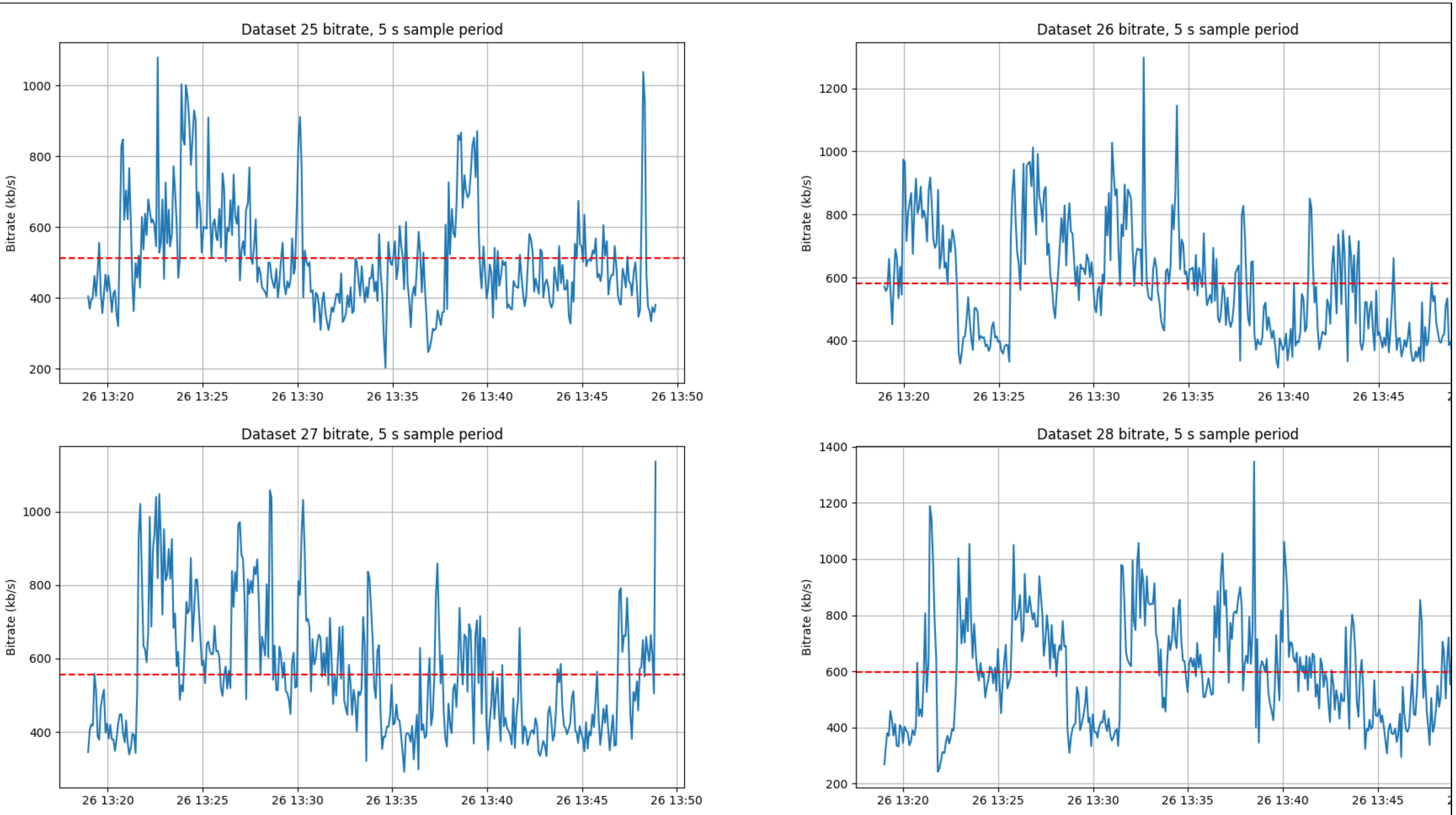


Fig. 71: Bitrate measured at receiver's end, while receiving 40 video streams, streams 25 – 28. Dashed red line shows mean received bitrate.

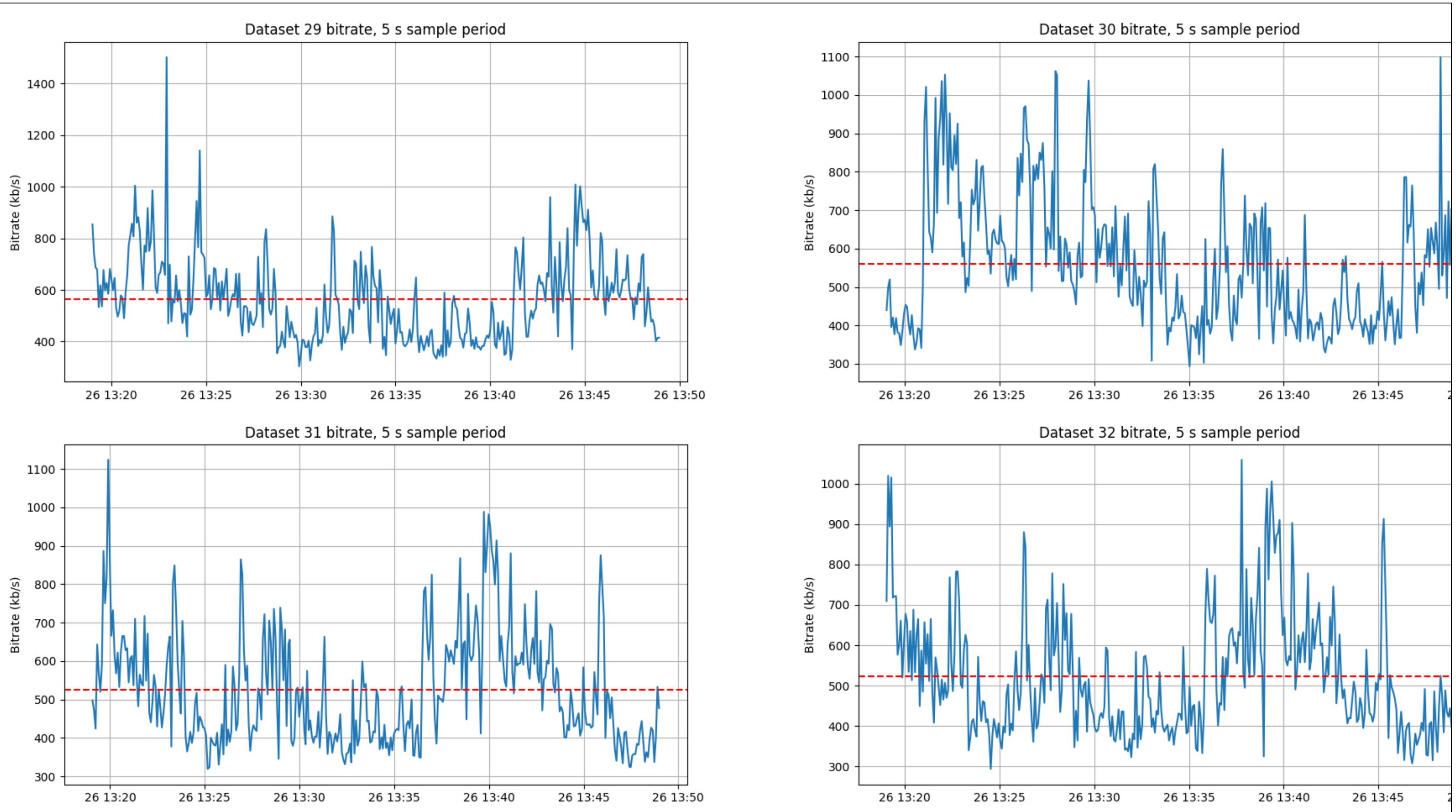


Fig. 72: Bitrate measured at receiver's end, while receiving 40 video streams, streams 29 – 32. Dashed red line shows mean received bitrate.

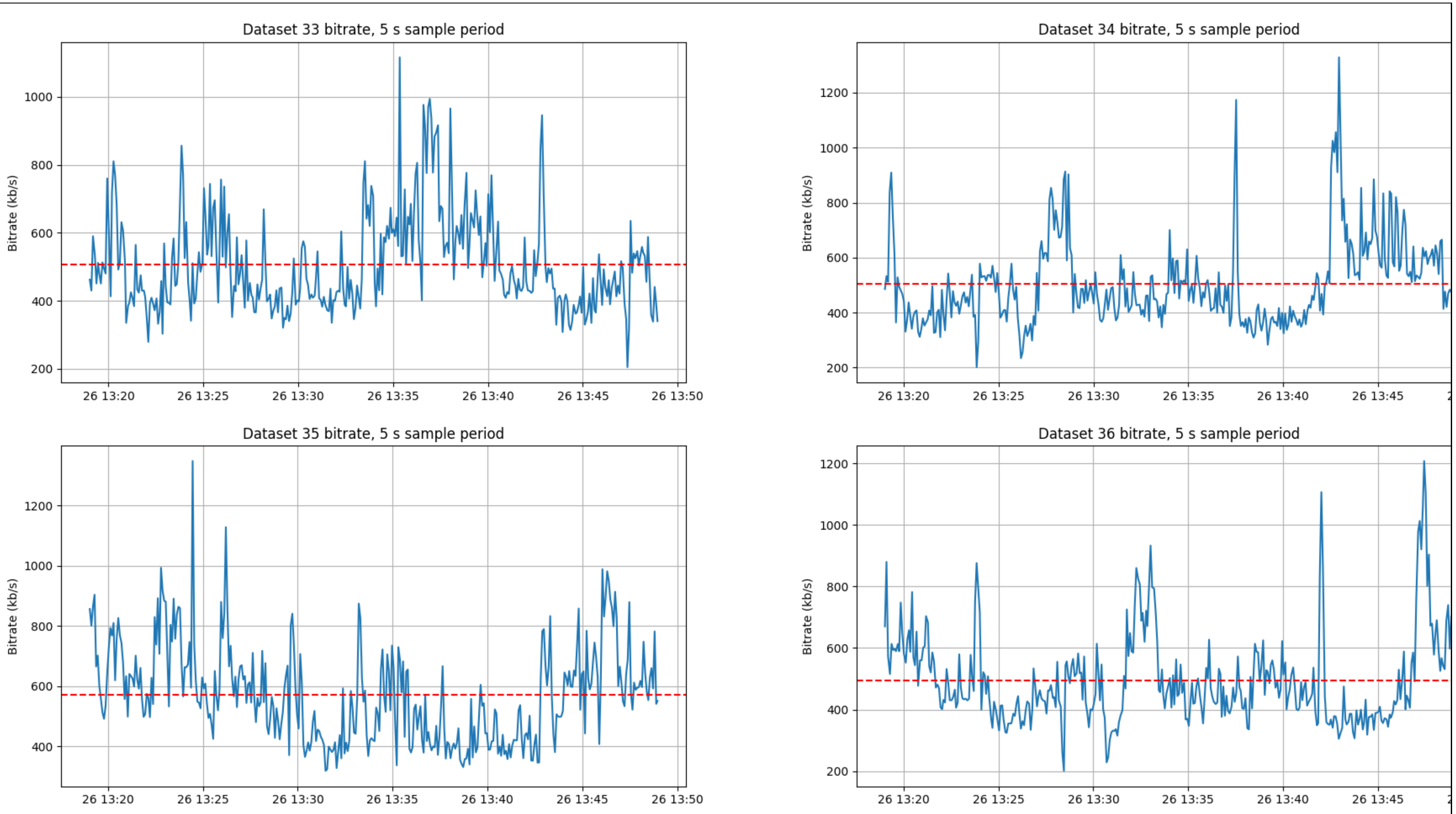
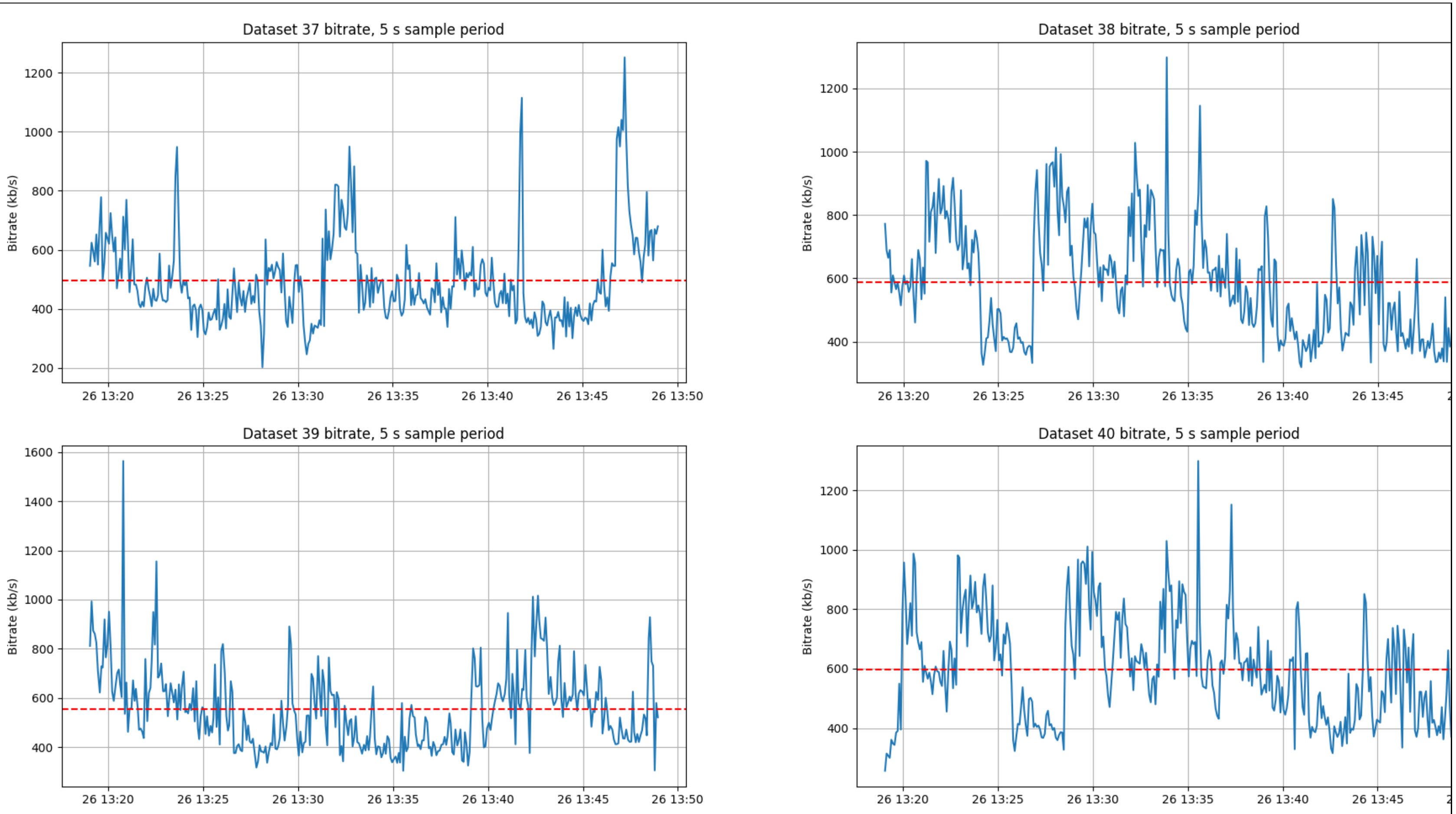


Fig. 73: Bitrate measured at receiver’s end, while receiving 40 video streams, streams 33 – 36. Dashed red line shows mean received bitrate.


Dataset 37 bitrate, 5 s sample period

Dataset 38 bitrate, 5 s sample period

Dataset 39 bitrate, 5 s sample period

Dataset 40 bitrate, 5 s sample period

Fig. 74: Bitrate measured at receiver’s end, while receiving 40 video streams, streams 37 – 40. Dashed red line shows mean received bitrate.

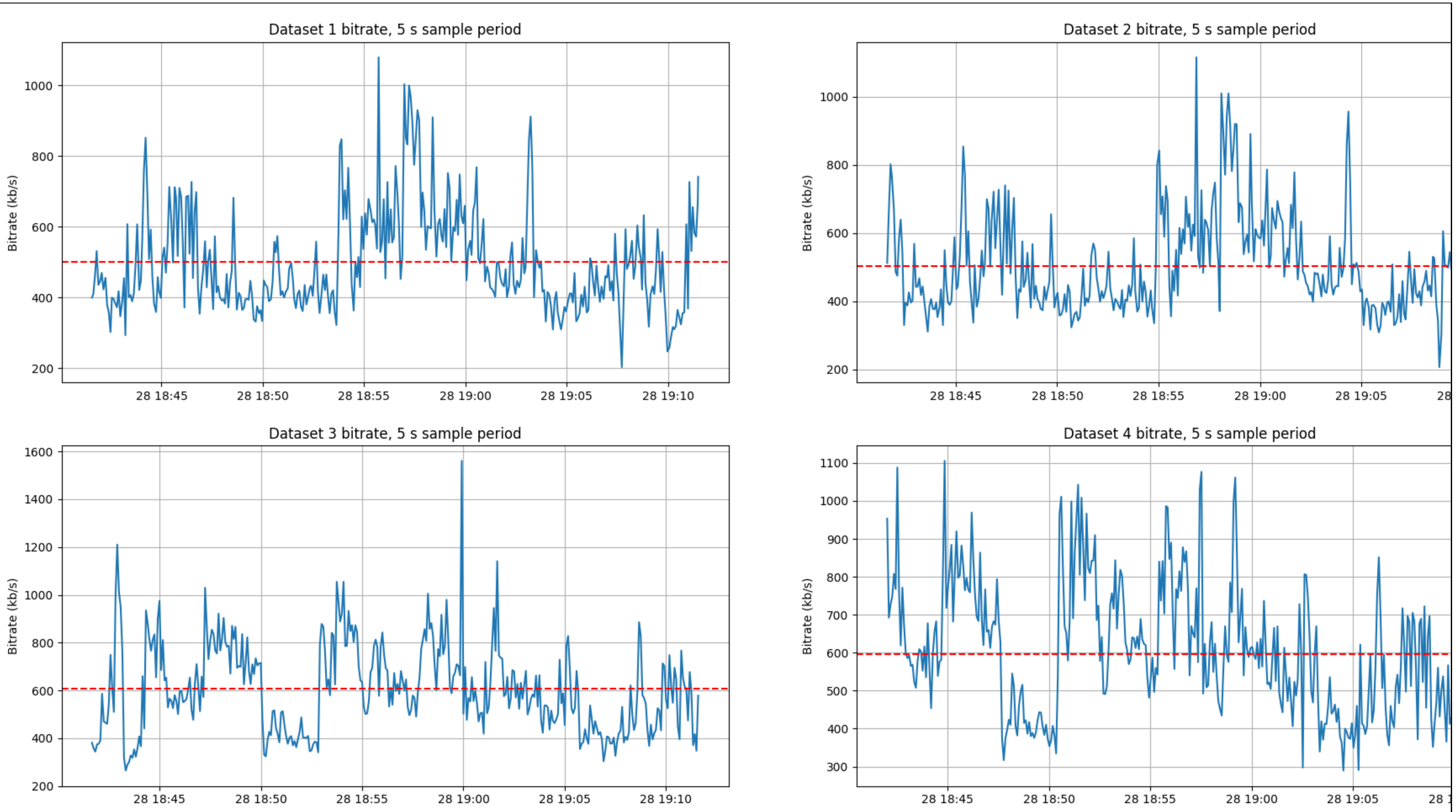


Fig. 75: Bitrate measured at receiver's end, while receiving 80 video streams, streams 1 – 4. Dashed red line shows mean received bitrate.

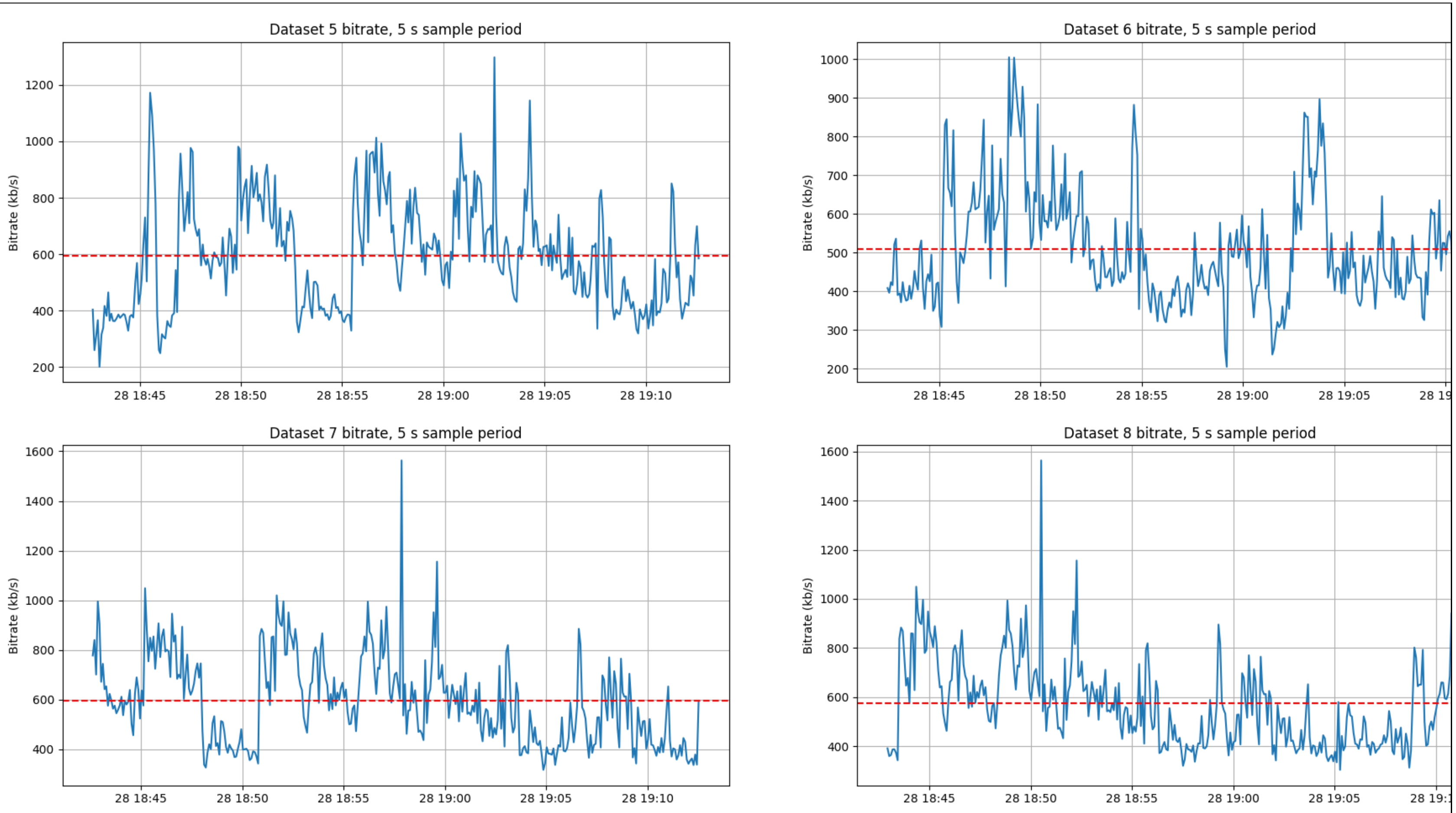


Fig. 76: Bitrate measured at receiver's end, while receiving 80 video streams, streams 5 – 7. Dashed red line shows mean received bitrate.

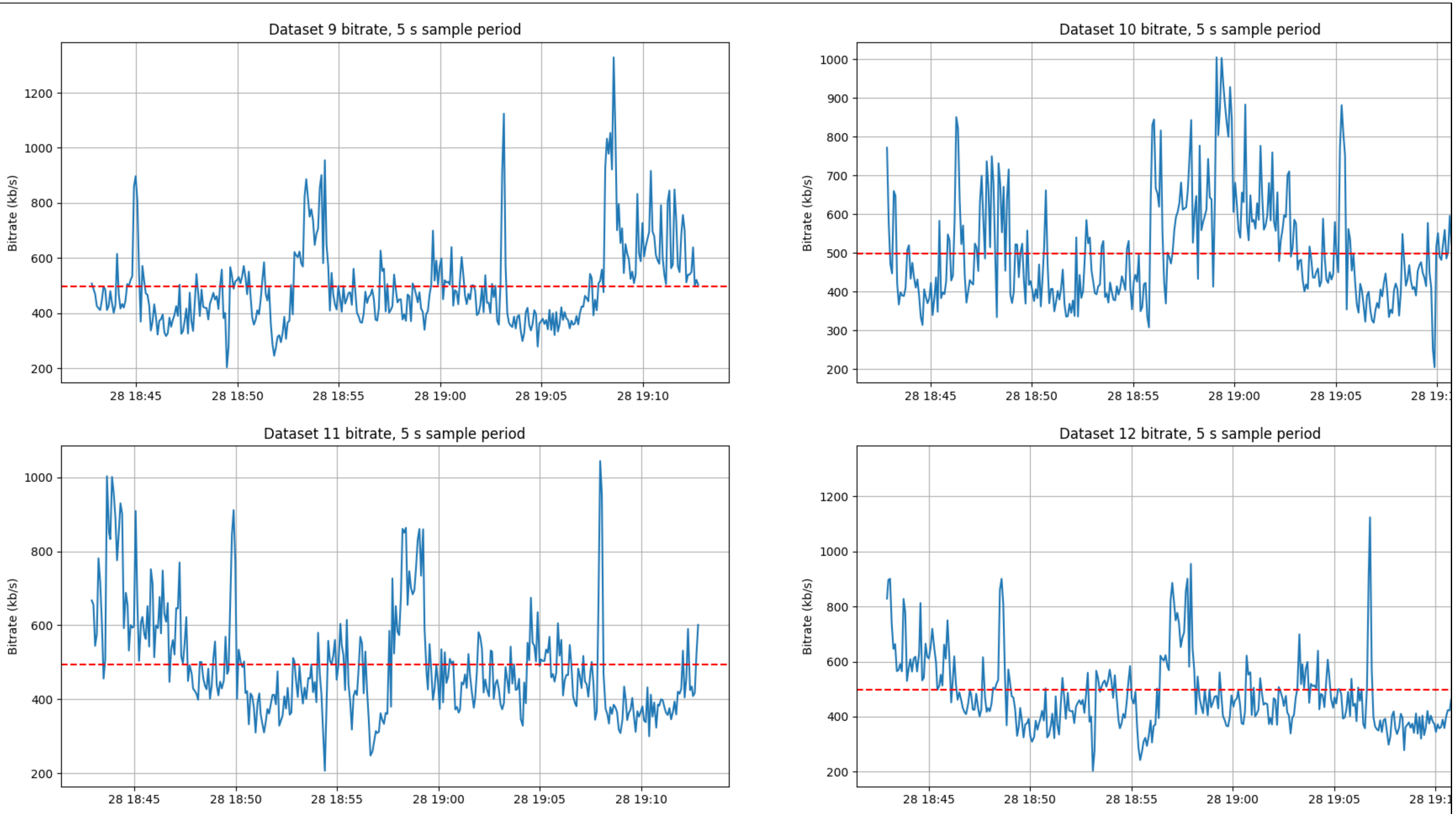


Fig. 77: Bitrate measured at receiver's end, while receiving 80 video streams, streams 9 – 12. Dashed red line shows mean received bitrate.

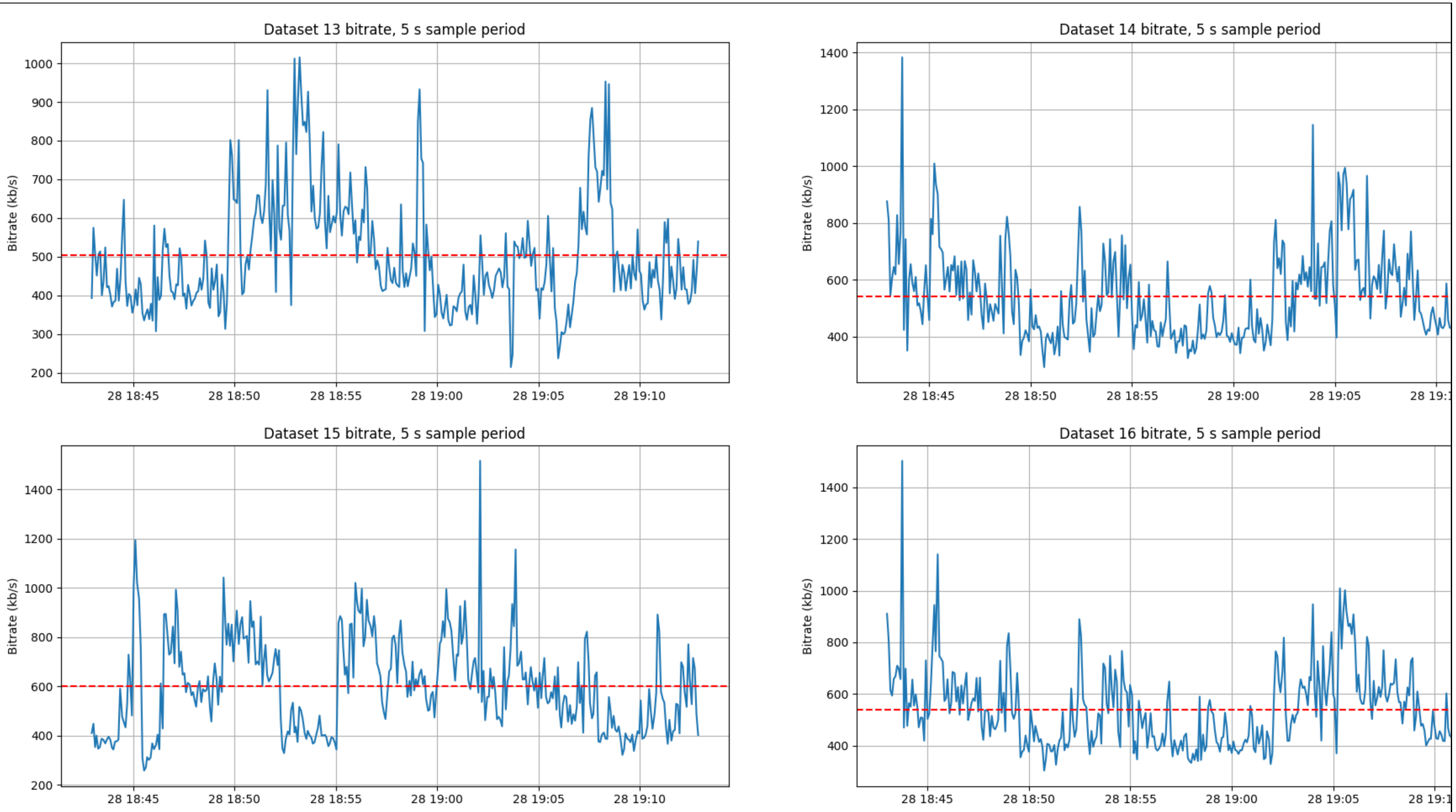


Fig. 78: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 13 – 16. Dashed red line shows mean received bitrate.

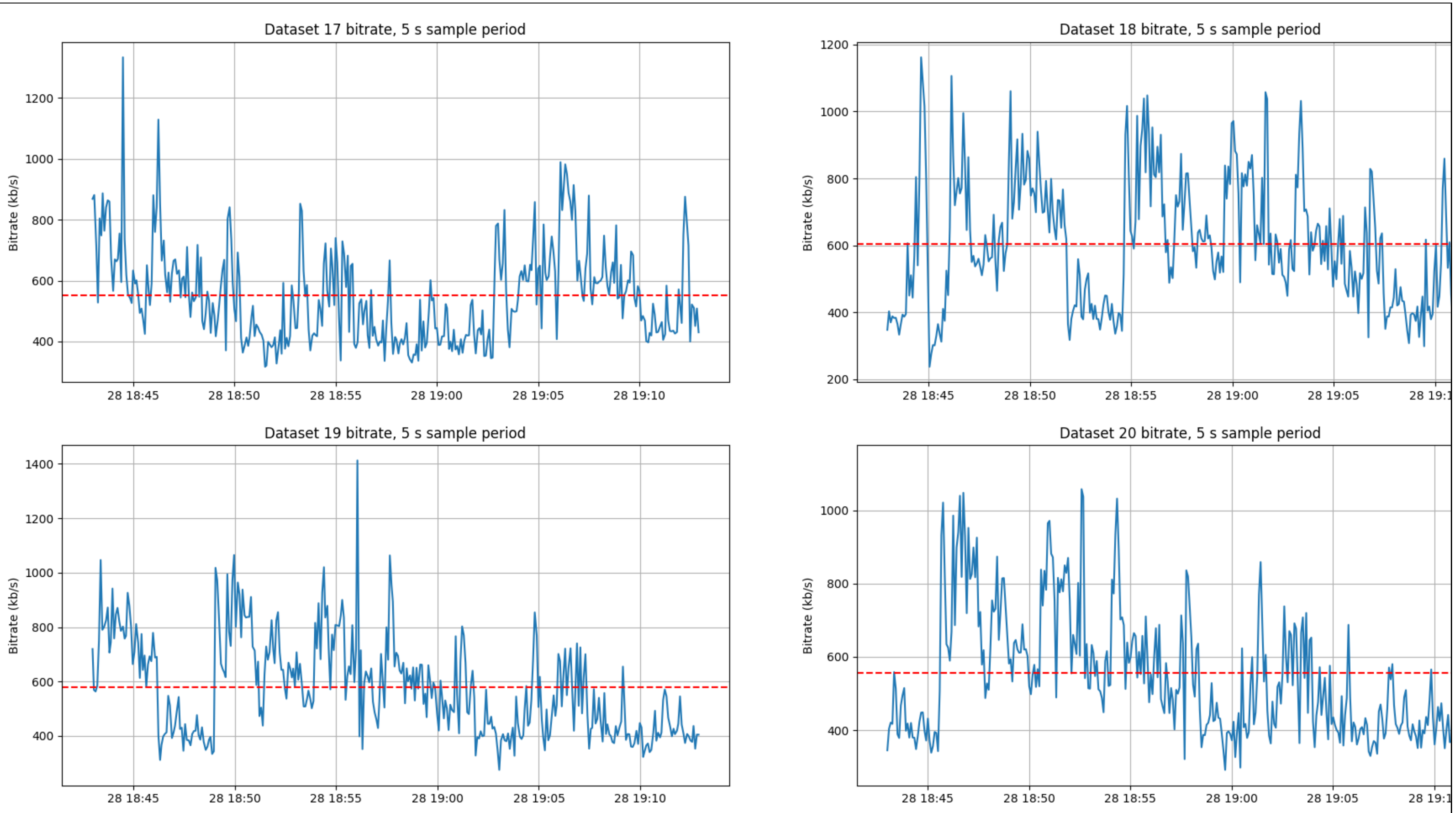


Fig. 79: Bitrate measured at receiver's end, while receiving 80 video streams, streams 17 – 20. Dashed red line shows mean received bitrate.

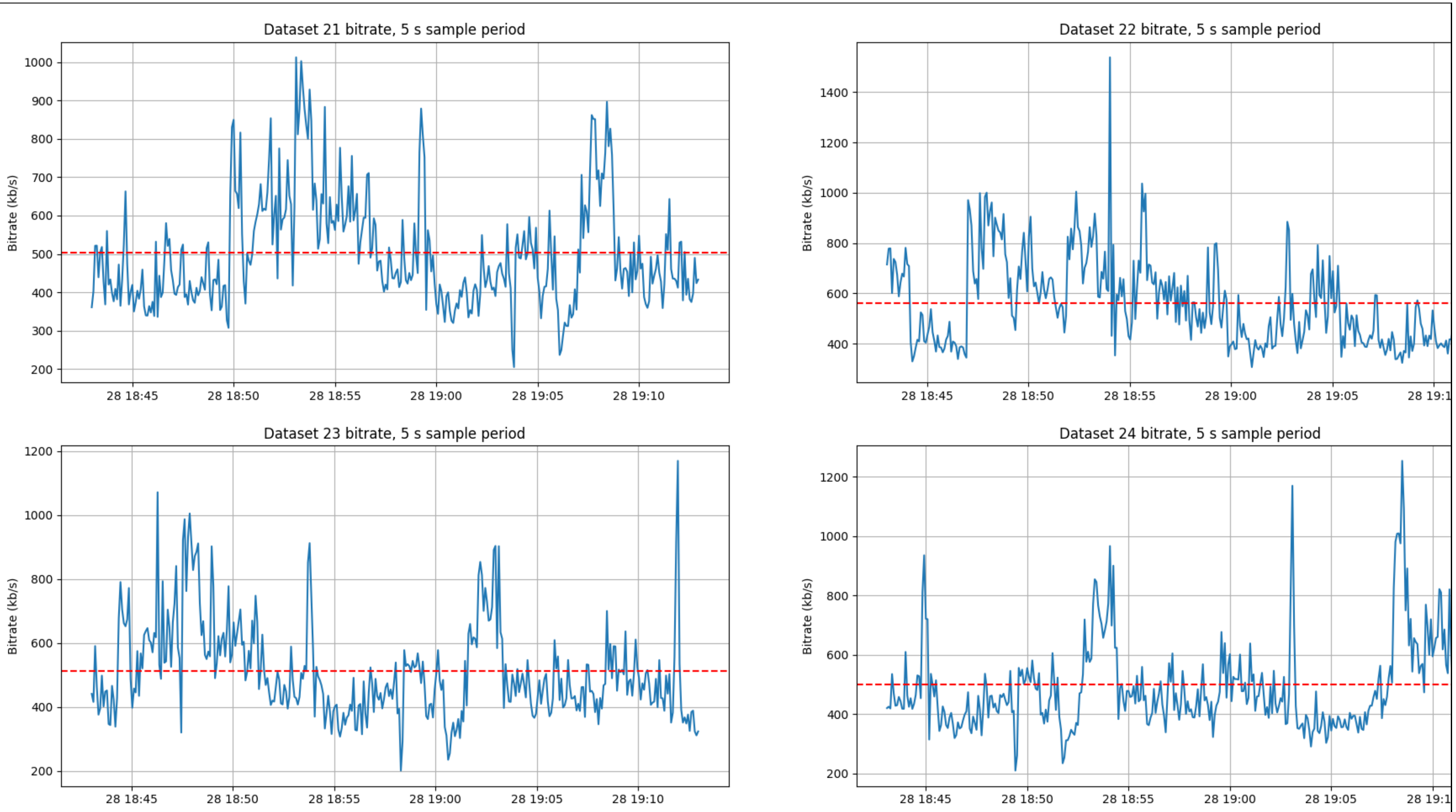


Fig. 80: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 21 – 24. Dashed red line shows mean received bitrate.

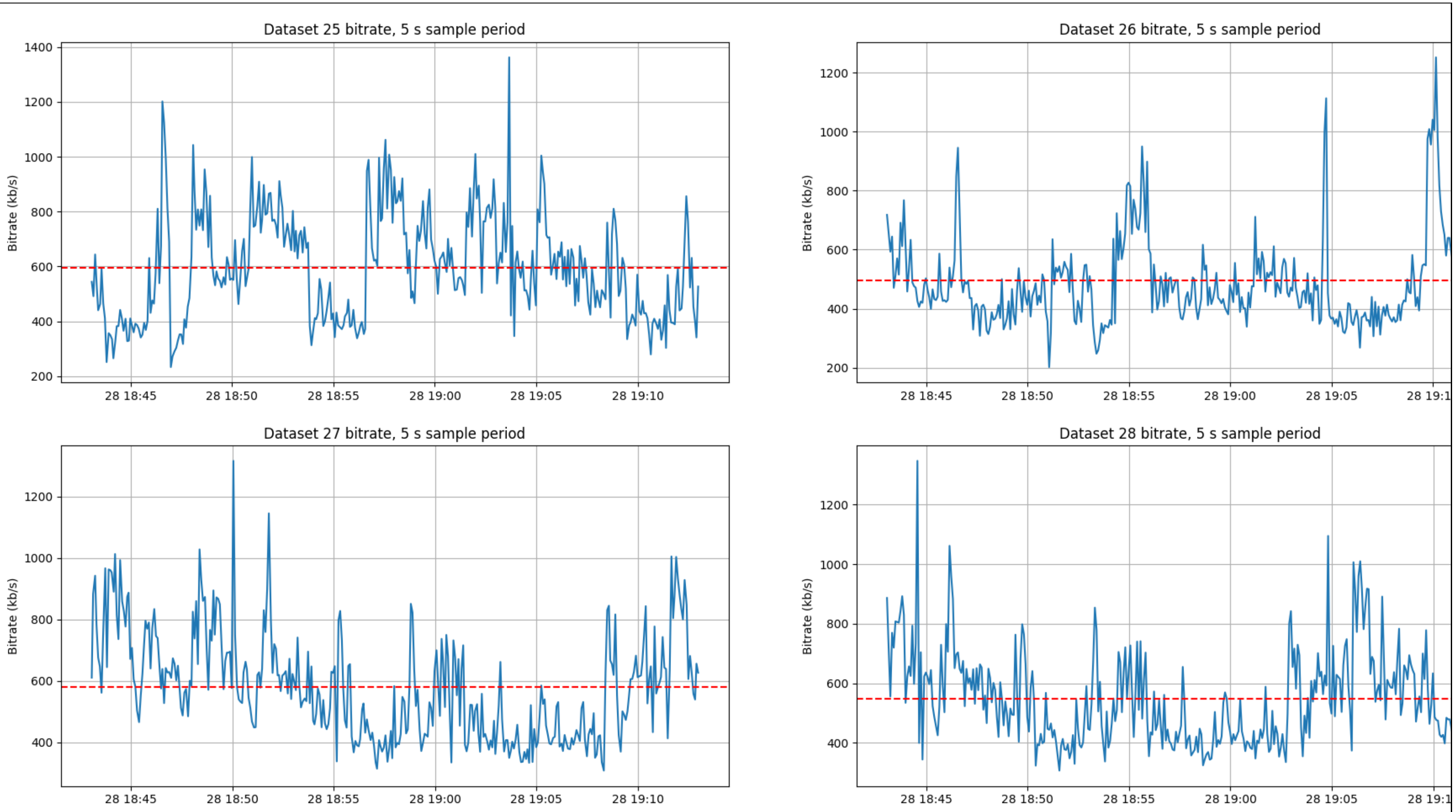


Fig. 81: Bitrate measured at receiver's end, while receiving 80 video streams, streams 25 – 28. Dashed red line shows mean received bitrate.

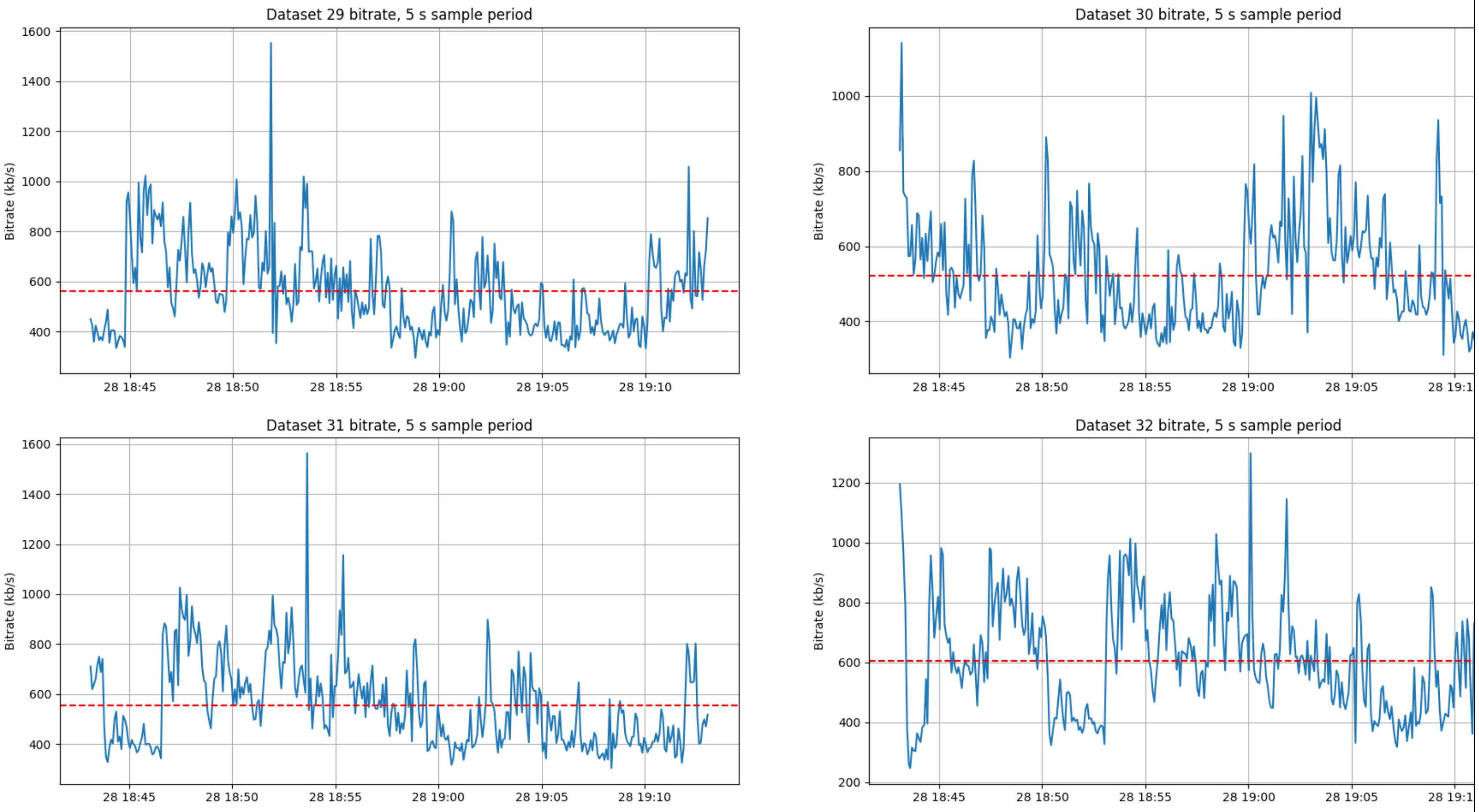


Fig. 82: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 29 – 32. Dashed red line shows mean received bitrate.

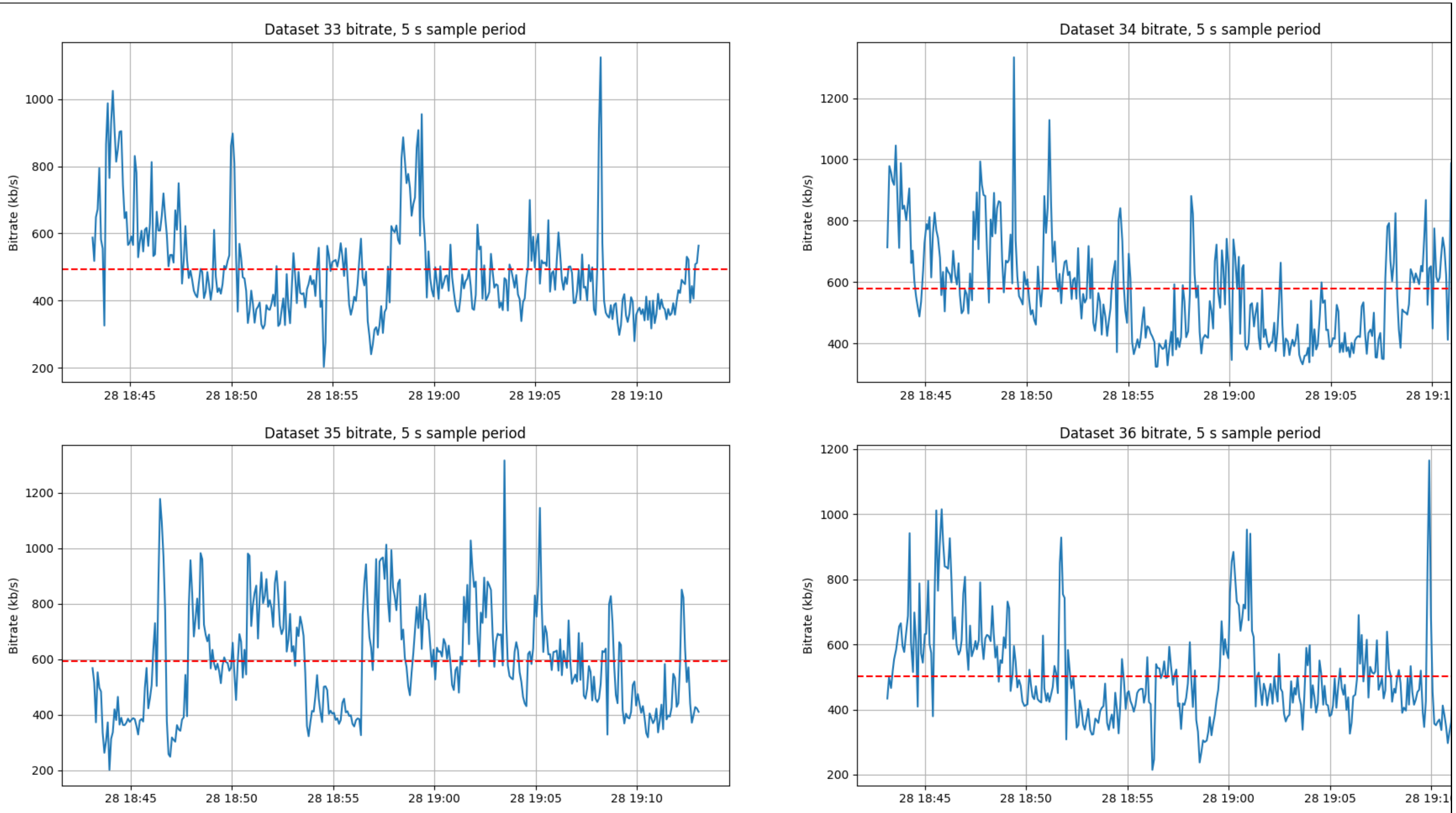


Fig. 83: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 33 – 36. Dashed red line shows mean received bitrate.

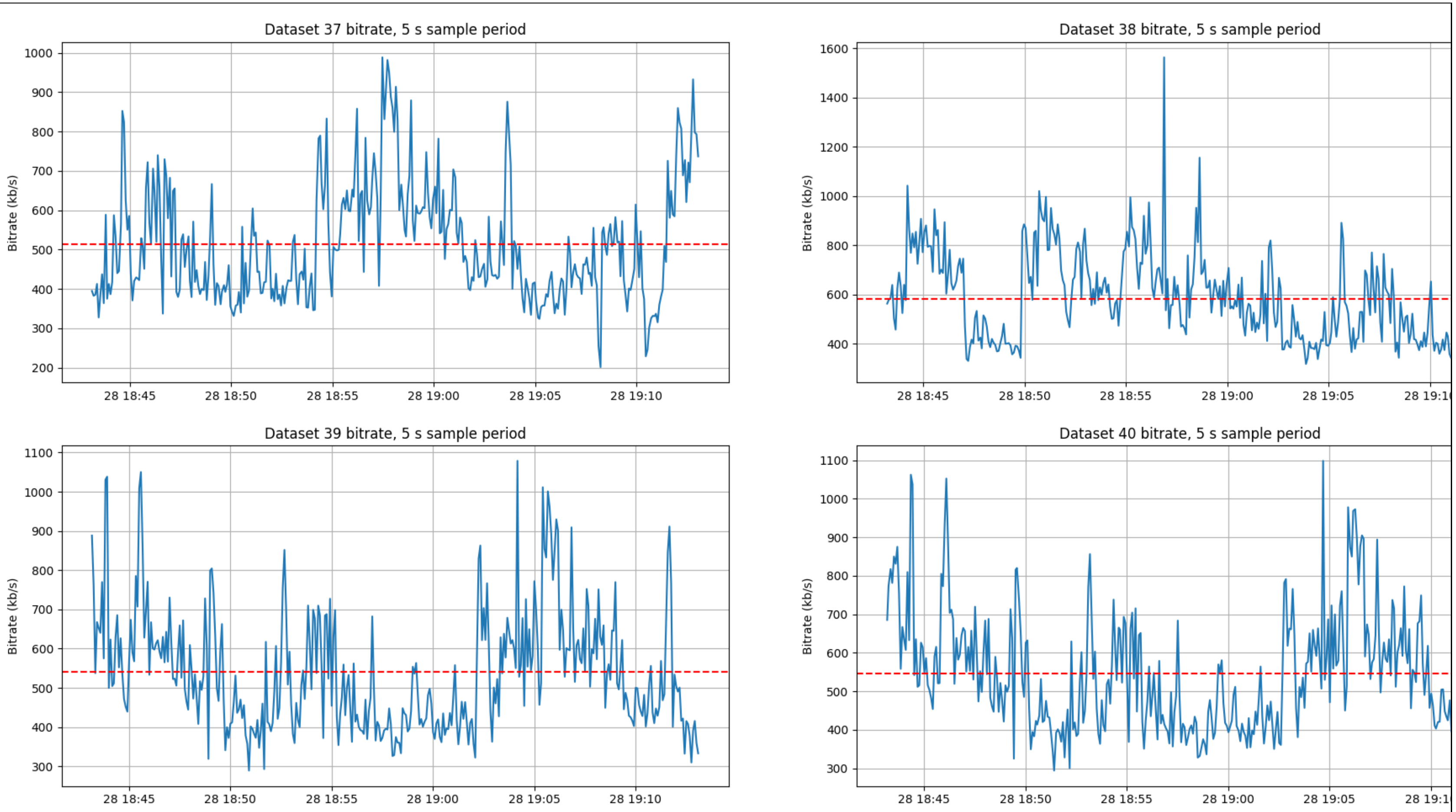


Fig. 84: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 37 – 40. Dashed red line shows mean received bitrate.

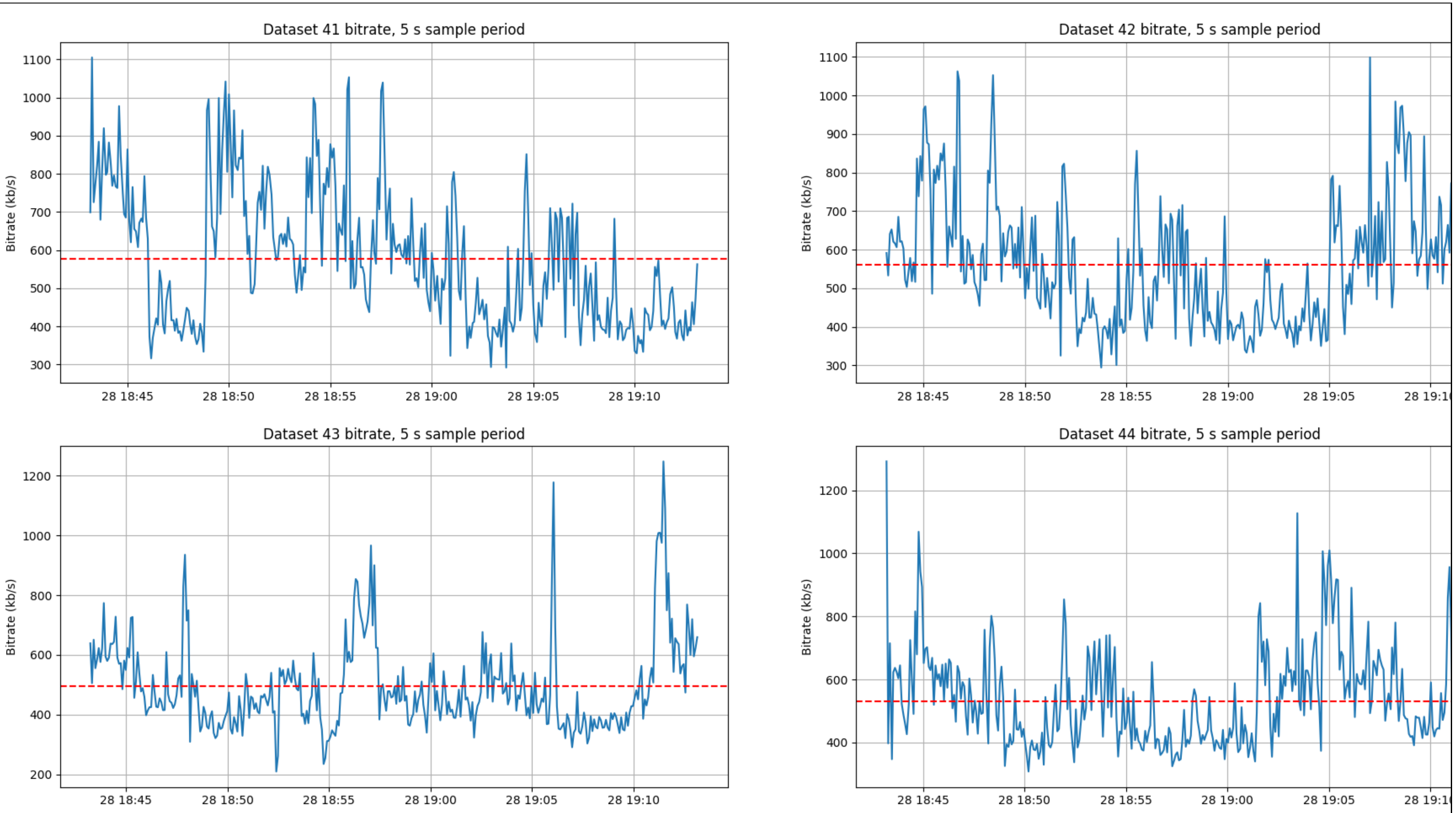
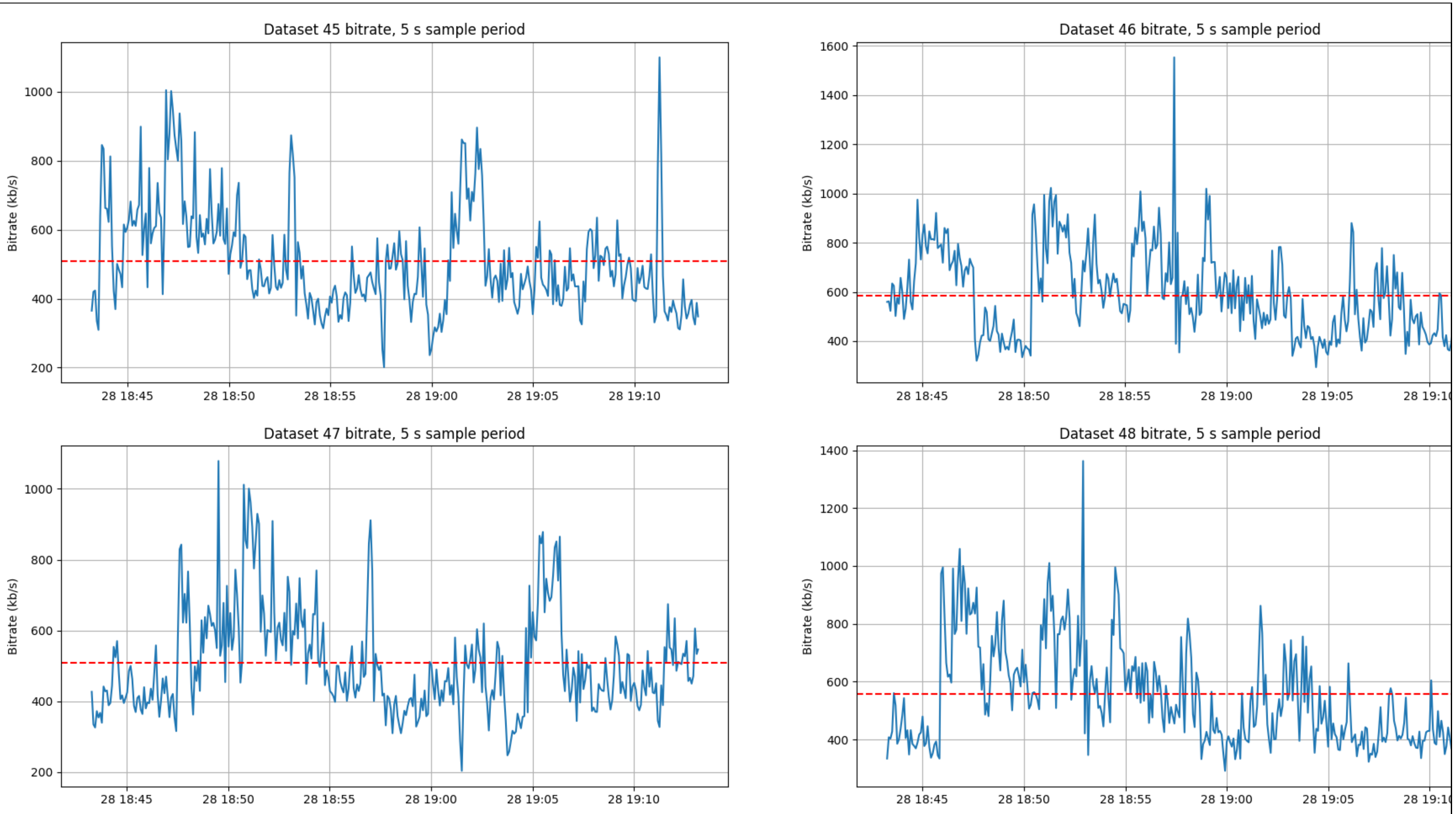


Fig. 85: Bitrate measured at receiver's end, while receiving 80 video streams, streams 41 – 44. Dashed red line shows mean received bitrate.


Dataset 45 bitrate, 5 s sample period

Dataset 46 bitrate, 5 s sample period

Dataset 47 bitrate, 5 s sample period

Dataset 48 bitrate, 5 s sample period

Fig. 86: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 45 – 48. Dashed red line shows mean received bitrate.

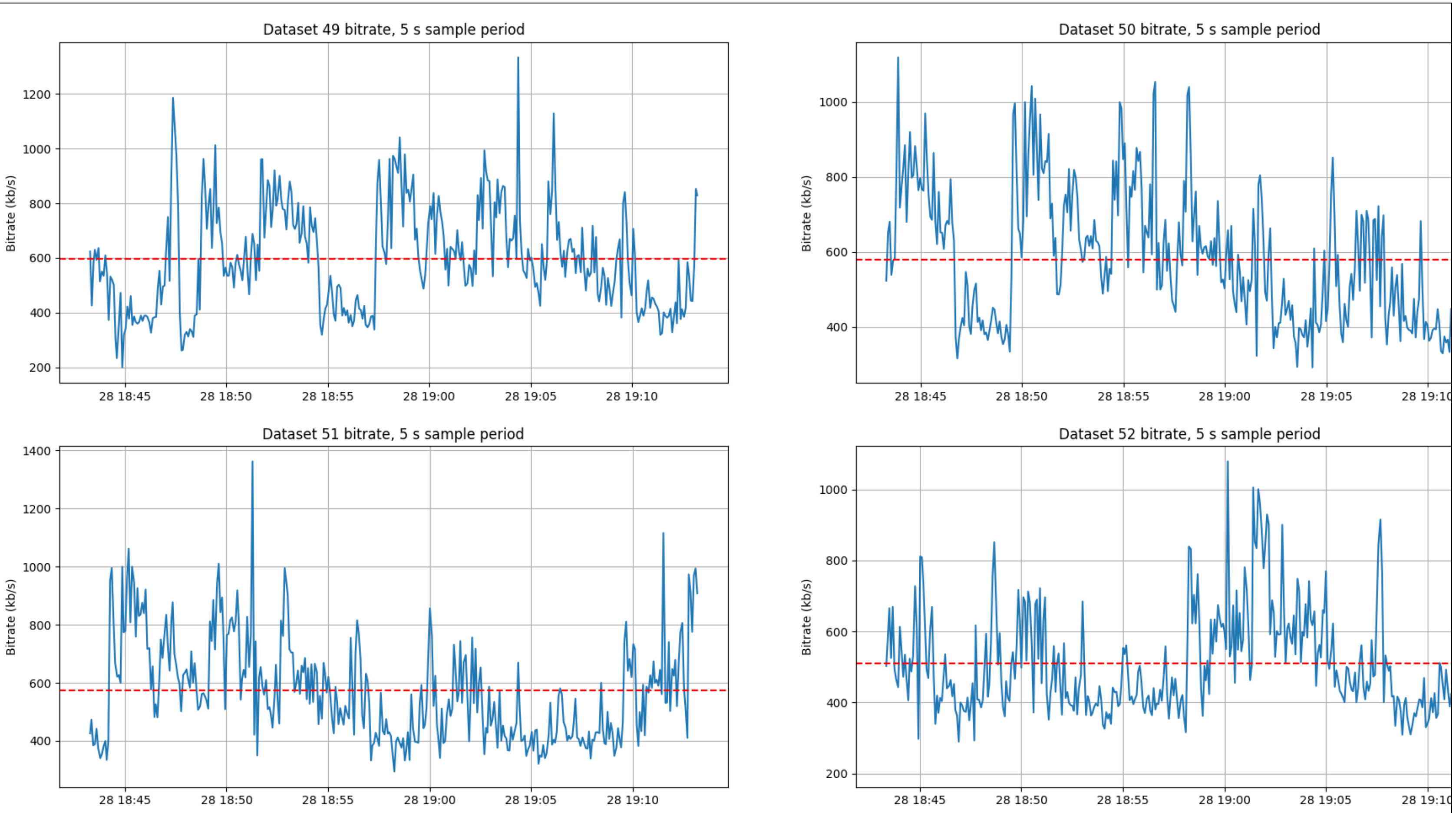
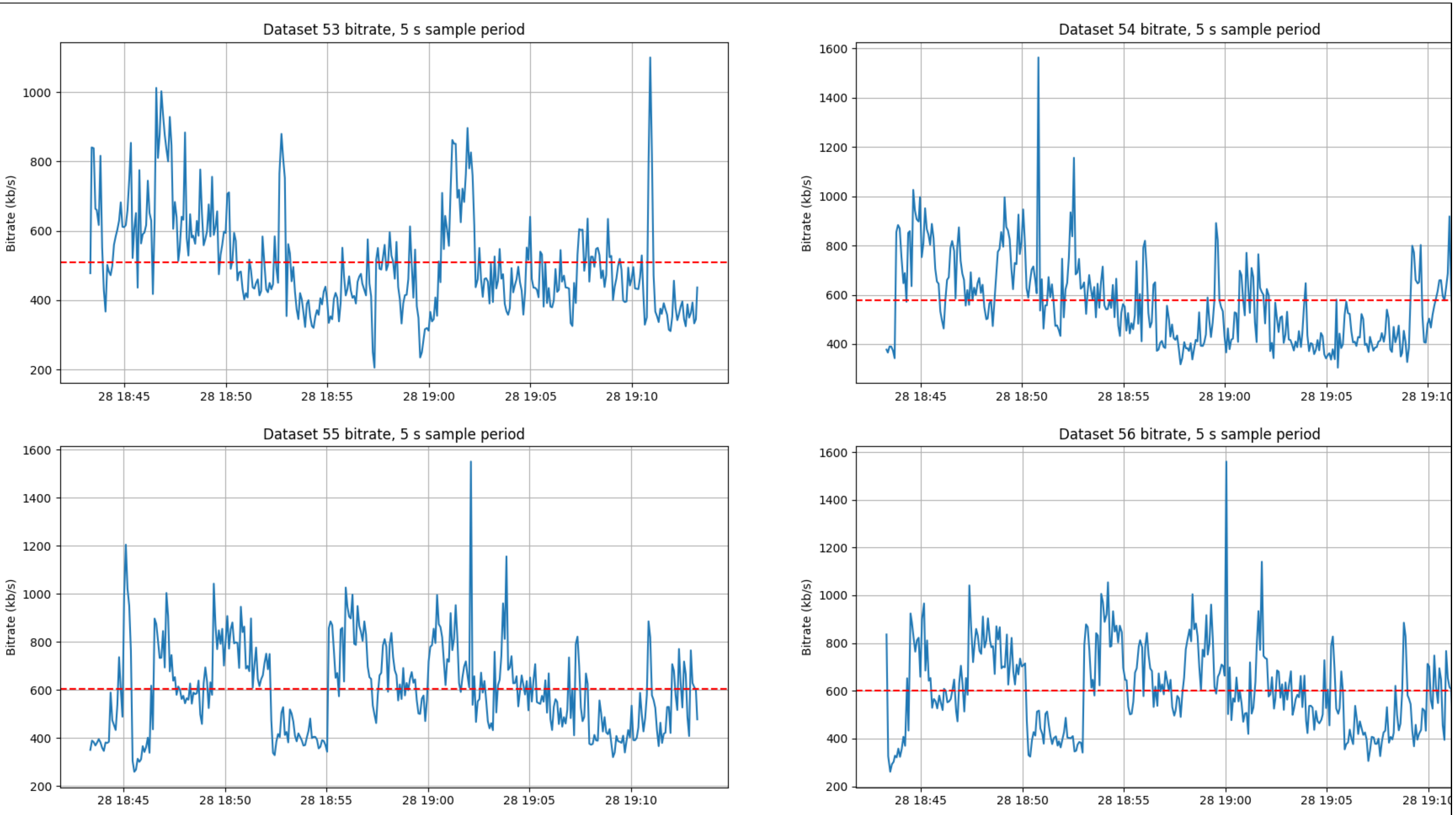


Fig. 87: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 49 – 52. Dashed red line shows mean received bitrate.


Dataset 53 bitrate, 5 s sample period

Dataset 54 bitrate, 5 s sample period

Dataset 55 bitrate, 5 s sample period

Dataset 56 bitrate, 5 s sample period

Fig. 88: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 53 – 56. Dashed red line shows mean received bitrate.

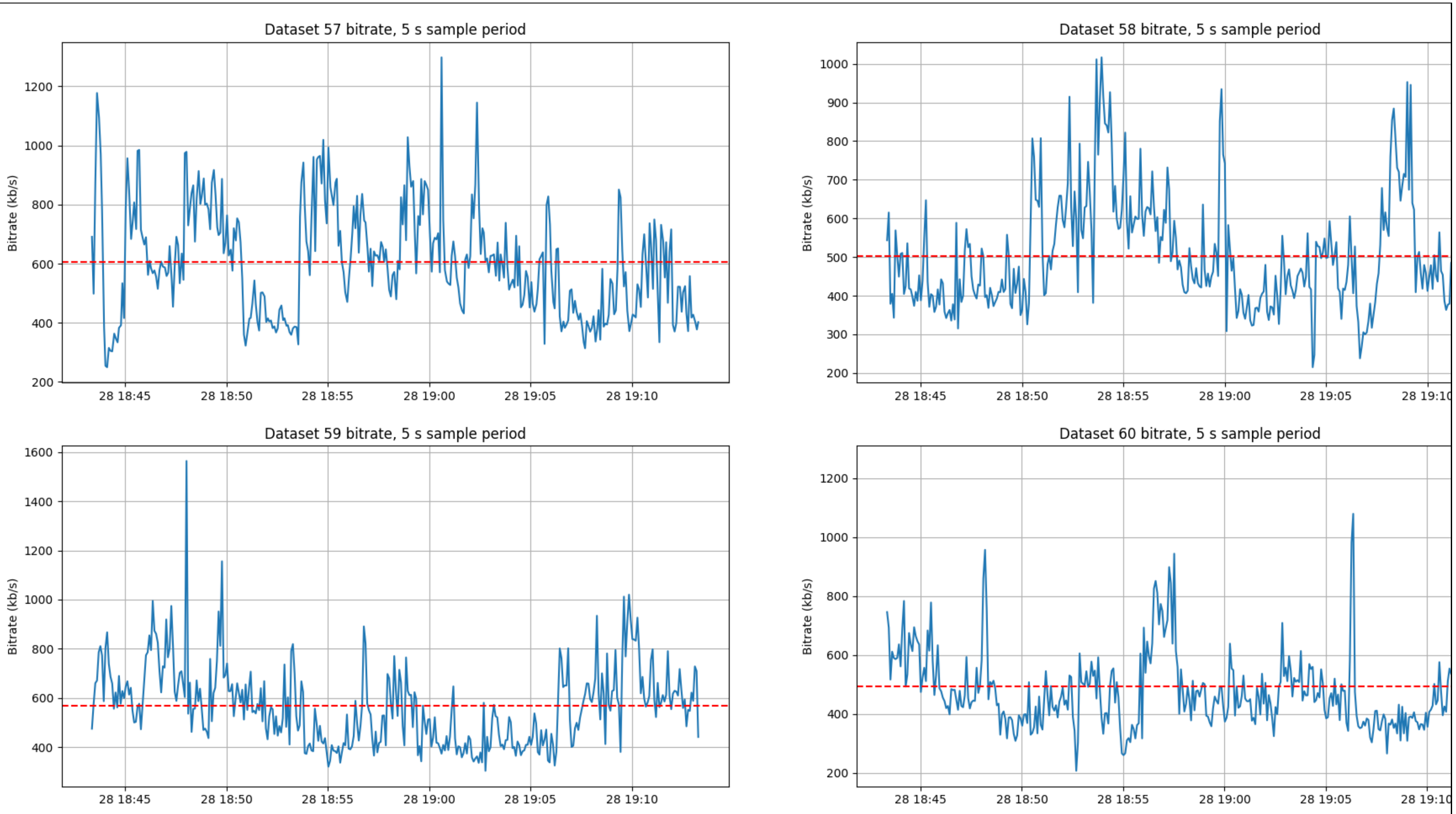
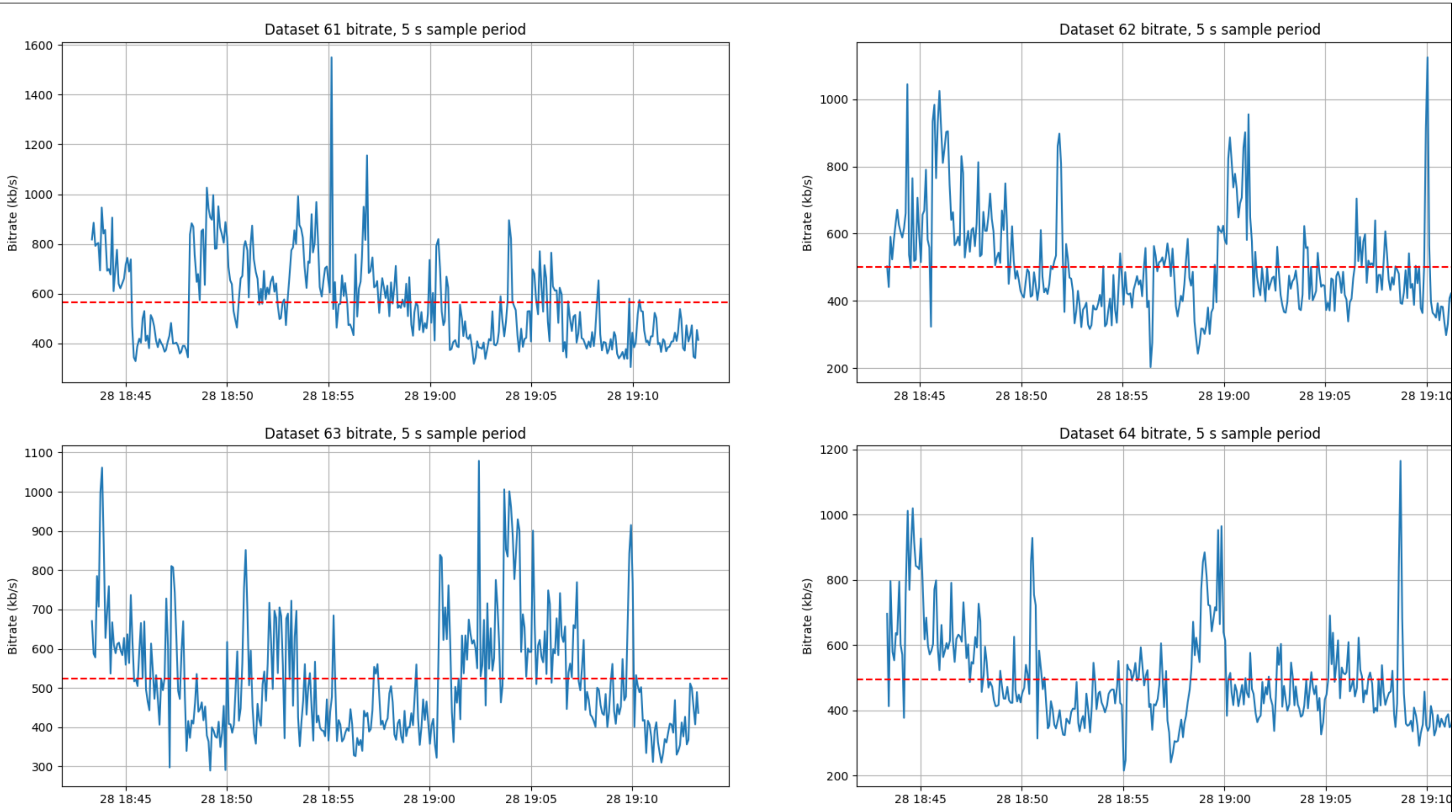


Fig. 89: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 57 – 60. Dashed red line shows mean received bitrate.



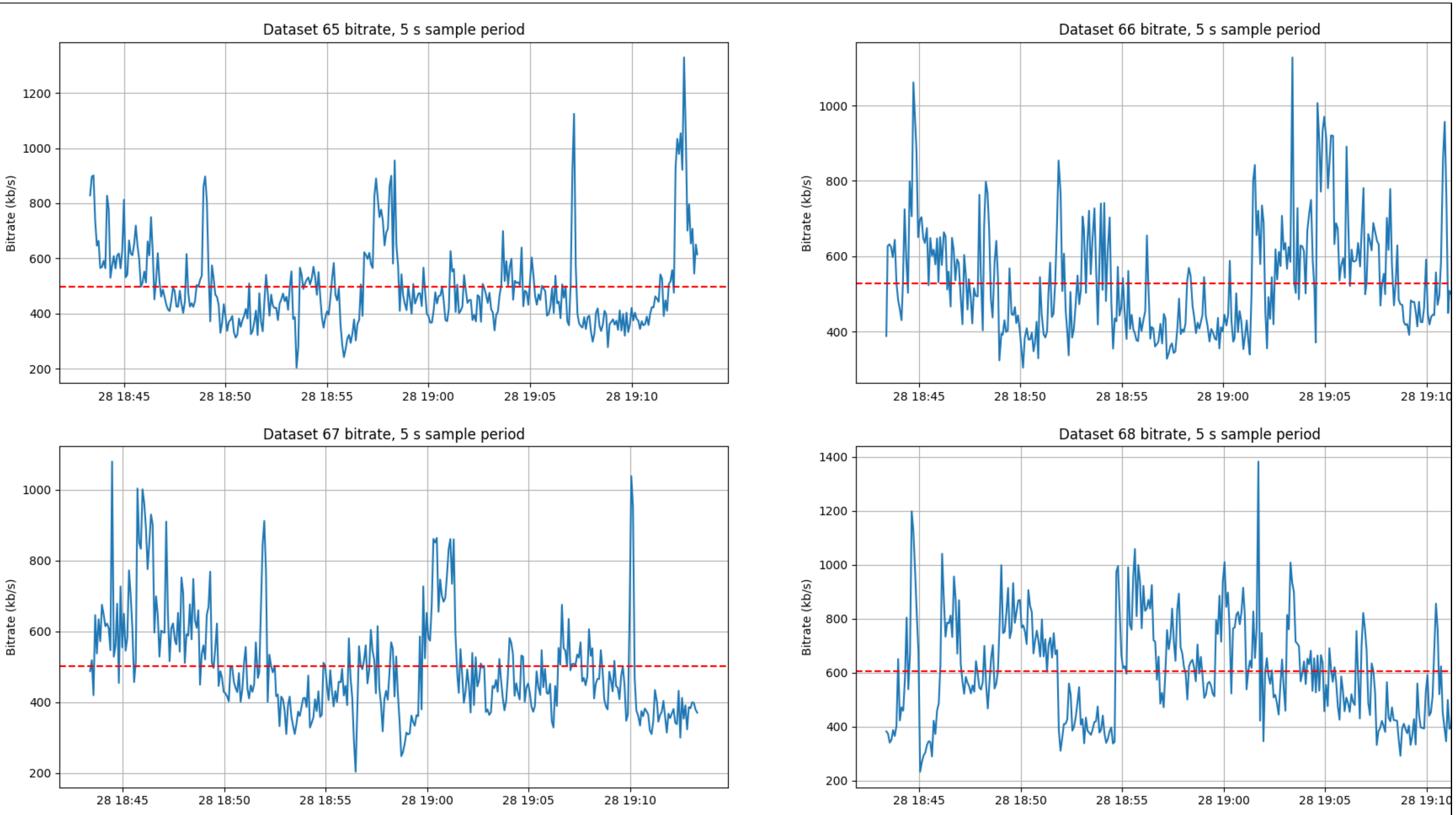
Dataset 61 bitrate, 5 s sample period

Dataset 62 bitrate, 5 s sample period

Dataset 63 bitrate, 5 s sample period

Dataset 64 bitrate, 5 s sample period

Fig. 90: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 61 – 64. Dashed red line shows mean received bitrate.

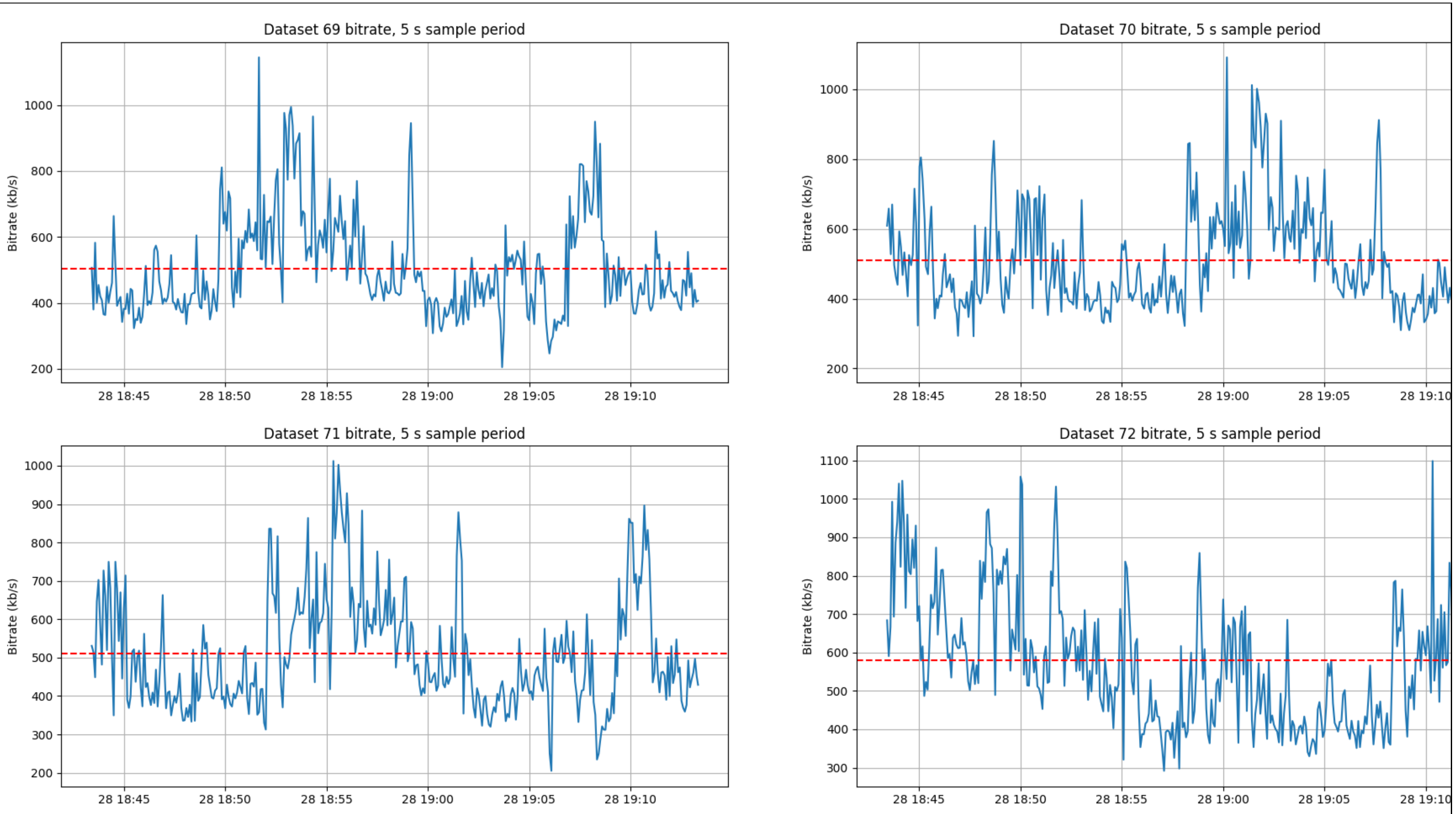

Dataset 65 bitrate, 5 s sample period

Dataset 66 bitrate, 5 s sample period

Dataset 67 bitrate, 5 s sample period

Dataset 68 bitrate, 5 s sample period

Fig. 91: Bitrate measured at receiver's end, while receiving 80 video streams, streams 65 – 68. Dashed red line shows mean received bitrate.



Dataset 69 bitrate, 5 s sample period

Dataset 70 bitrate, 5 s sample period

Dataset 71 bitrate, 5 s sample period

Dataset 72 bitrate, 5 s sample period

Fig. 92: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 69 – 72. Dashed red line shows mean received bitrate.

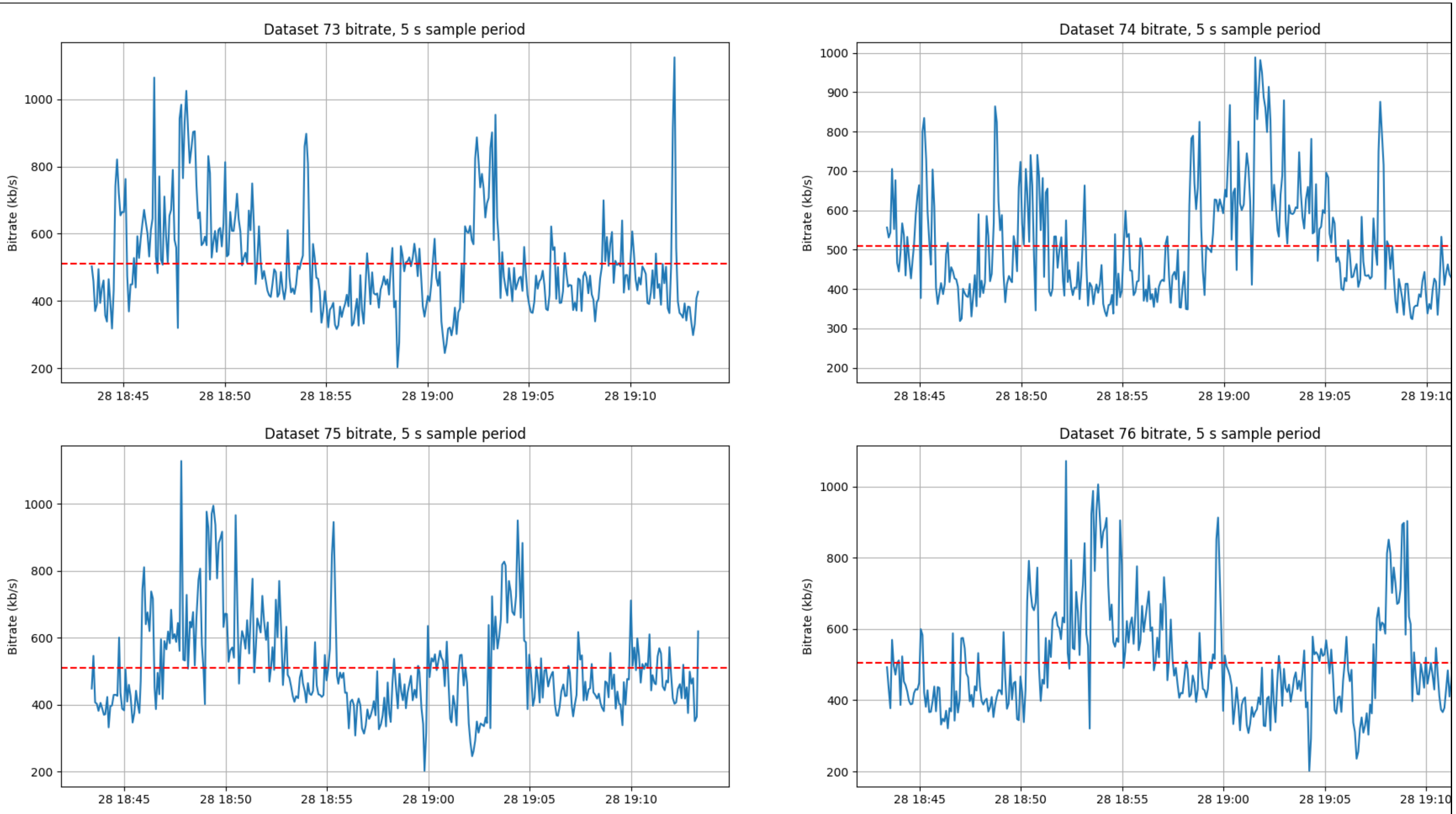


Fig. 93: Bitrate measured at receiver's end, while receiving 80 video streams, streams 72 – 76. Dashed red line shows mean received bitrate.

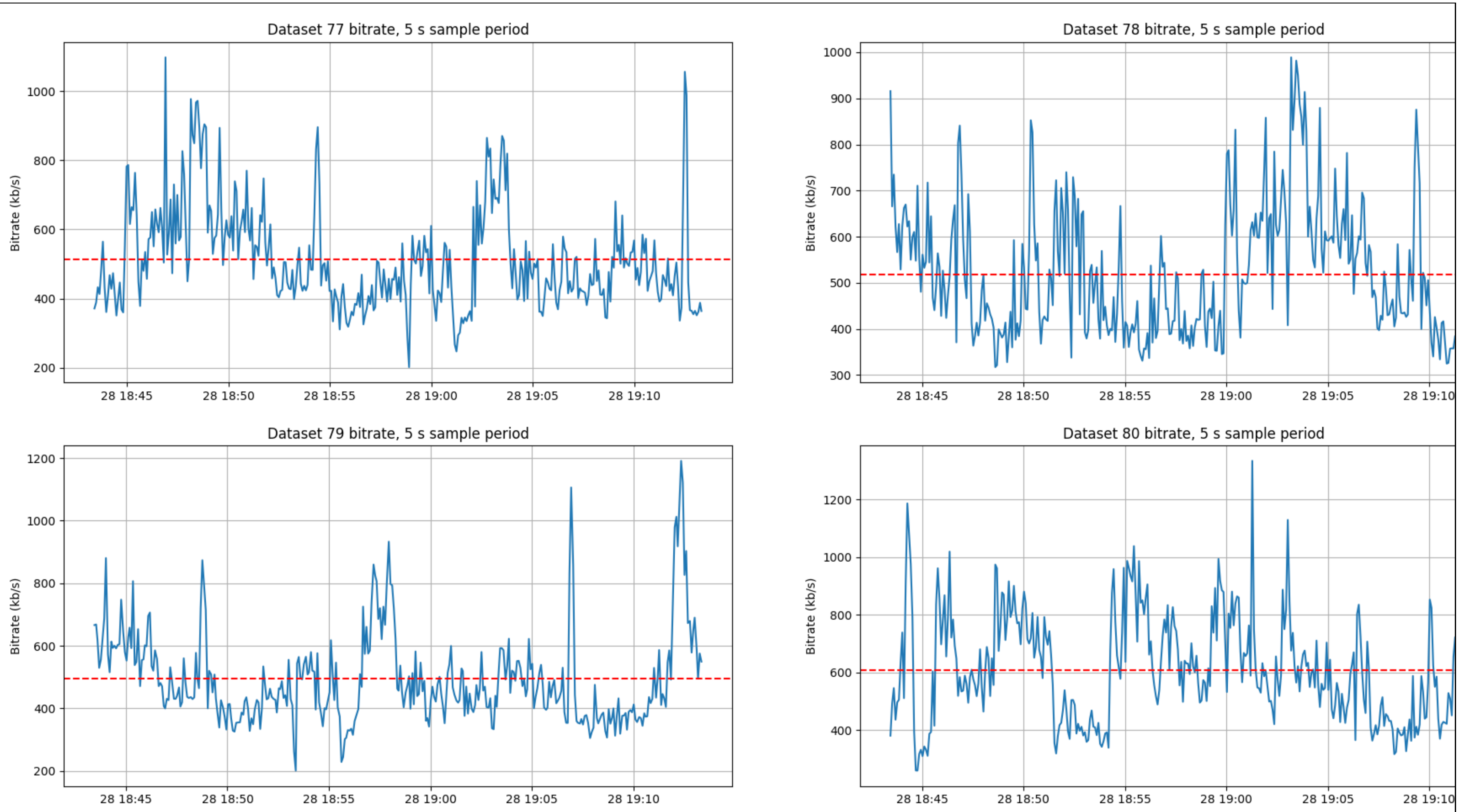


Fig. 94: Bitrate measured at receiver’s end, while receiving 80 video streams, streams 76 – 80. Dashed red line shows mean received bitrate.

Appendix 9: Ansible playbooks

Listing 1: one_containerized_stream.yml (called by orchestrator for every containerized stream)

```
---
- name: Create a streaming container
  hosts: videoserver
  become: no
  tasks:
  - name: Create the container
    shell: "docker run -d -e PORT={{ streaming_port }} -e VIDEO={{
video_file }} -e START_AT={{ start_at }} -e VIDEO_DURATION={{
video_duration }} --network host -v /mnt/ramdisk:/videos --name {{
cont_name }} video_streamer"

- name: Videoclient on condominium requests the stream
  hosts: condominium
  become: no
  tasks:
  - name: Create the working directory
    file:
      path: /home/videoclient/bitrate_logs/{{ cont_name }}
      state: directory
      mode: '0755'
  - name: Request the stream
    shell: "tsp -I http http://10.0.0.1:{{ streaming_port }} -O drop -P
analyze --interval 5 --multiple-files --ts-analysis --output-file testing"
    args:
      chdir: /home/videoclient/bitrate_logs/{{ cont_name }}

- name: Destroy the container
  hosts: videoserver
  become: no
  tasks:
  - name: Remove the container created previously
          shell: "docker rm {{ cont_name }}"
```

Listing 2: one_native_stream.yml (called by orchestrator for every native stream)

```
---
- name: Prepare a listening streamer
  hosts: videoserver
  become: no
  tasks:
  - name: Start socat listener
    shell: >
      socat TCP-LISTEN:{{ streaming_port }},reuseaddr,fork SYSTEM:'sh /home/videoserver/native/stream.sh {{ video_file }} {{ start_at }} {{ video_duration }}'
    args:
      chdir: /mnt/ramdisk
    async: 3600
    poll: 0

- name: Videoclient on condominium requests the stream
  hosts: condominium
  become: no
  tasks:
  - name: Create the logging directory
    file:
      path: /home/videoclient/bitrate_logs/native/{{ number_of_streams }}/{{ streaming_port }}
      state: directory
      mode: '0755'
  - name: Request the stream
    shell: "tsp -I http http://10.0.0.1:{{ streaming_port }} -O drop -P analyze --interval 5 --multiple-files --ts-analysis --output-file testing"
    args:
      chdir: "/home/videoclient/bitrate_logs/native/{{ number_of_streams }}/{{ streaming_port }}"
```

Listing 3: prelude.yml (called by orchestrator as a prelude to one_native_stream.yml)

```
---
- name: Start the power meter
  hosts: videoserver
  become: yes
  tasks:
  - name: Start powertop
    shell:    /home/videoserver/ansible_documents/start_powertop.sh    {{
iterations }}

- name: Create a directory for bitrate logs on videoclient
  hosts: condominium
  become: no
  tasks:
  - name: Create the parent directory for native streaming bitrate logs
    file:
      path: /home/videoclient/bitrate_logs/native/{{ number_of_streams }}
      state: directory
      mode: '0755'}}"
```

Listing 4: powertop.yml (called by orchestrator as a prelude to one_containerized_stream.yml)

```
---
- name: Start the power meter
  hosts: videoserver
  become: yes
  tasks:
  - name: Start powertop
    shell:    /home/videoserver/ansible_documents/start_powertop.sh    {{
iterations }}
```