Overview of Peer-to-Peer networks and challenges in trust and reputation

Andrea Mangion, Mark Micallef
Department of Computer Science
Faculty of Information and Communication Technology
University of Malta
aman0001@um.edu.mt

***Abstract*—The use of peer-to-peer networks is becoming a more popular way how to exchange content over a network. An increase in use of peer-to-peer networks does not only bring along advantages, but also challenges. In this paper we give an overview of peer-to-peer networks and why they are more attractive when compared to the traditional client/server approach. Trust and reputation are two of the main challenges that are faced in peer-to-peer setups. However, there are a number of approaches which tackle such issues.**

Keywords— peer-to-peer, incentive, trust, security, reputation

I. INTRODUCTION

The birth of P2P (peer-to-peer) networks is considered by many to have taken place in 1999 with the creation of Napster (mainly used for sharing music), Freenet (sharing documents in anonymous manner) and SETI@home (volunteer computing - computers connected to the Internet offer their resources to process radio data in finding extra terrestrial intelligence). Unfortunately, there were instances were peer-to-peer networks were associated to illegalities and misuse. Napster was eventually used in order to share pirated music and this and other unresolved legal issues have led to the closure of Napster. However, P2P computing is now being used in different sectors other than sharing music. BitTorrent is used to distribute updates to software and files containing media content to a large number of nodes [2]. Organisations can use commercial peer-to-peer solutions allowing them to spread news and information to their clients and employees [3]. Skype is also considered to be a modern adaptation of peer-to-peer networks, used by millions of people around the globe to make video and phone calls [4].

One of the main and important characteristics in a P2P network is that there is no central server. Every node connected to a pure P2P network can be either a client or a server. A peer requiring a particular resource will normally broadcast such request to the network. Peers that have the resource available will provide the requesting peer with the data required. Depending on the protocol applied, such activity may differ. For example, a number of peers can contribute a part of the required resource, rather than the whole of it. A requester may choose not to send the query to all peers in the network but to a selected group of peers depending on their availability and reputation.

Having a p2p network with zero degree of centralization makes it more even more difficult to control peers who have bad intentions. Among the heterogeneous peers, some might be honest and provide high-quality services, some might be self-serving and not want to provide services for other peers, some might be even malicious by providing bad services or harming the consumers [9].

Here I will give an overview of what the paper is going to cover in sections. Mainly the first section will be on p2p networks, why they are good etc., then the proposals for ensuring trust and reputation, any results and compare them and then the conclusion

II. BACKGROUND

The best features that are highlighted in a p2p network are:

High degree of decentralization. This is the heart of a pure p2p network. Peers can be either client or servers. There are p2p structures that allow certain super peers to have more of a centralized role. The major number of nodes should have a dynamic state, to allow the highest degree of decentralization in the network.

Self-organization. It is easy for a node to join a p2p network. Normally, a node is part of a network once it provides information such as the IP address to a node that is already a member. Then, there should be no or very little manual intervention to configure and maintain the network.

Multiple administrative domains. No single organisation should own the nodes on a network. An individual should own a node(s) that choose to participate in the network.

Low barrier to deployment. Since nodes can easily join a p2p network, the cost of deployment is much lower than a conventional client-server setup.

Organic growth. The only way to upgrade a client-server setup is by replacing or adding to existing assets. A p2p network grows effortlessly since peers can join and contribute their resources.

Resilience to faults and attacks. High level of decentralization mean the resilience of p2p networks to faults and attacks is high. Only a few, if any, nodes should be critical to the network. For a p2p network to be attacked

or completely shut down, a large number of nodes have to be targeted.

Abundance and diversity of resources. A P2P network provides a large number of resources with combinations of environments located in various geographical sites. Such would be difficult or impossible as well as expensive for an organisation to replicate.

Among the heterogeneous peers, some might be honest and provide high-quality services, some might be self-serving and not want to provide services for other peers, some might be even malicious by providing bad services or harming the consumers [9]. Trust and reputation are vital in having a secure network. Trust is based on experiences that a peer has in relationship to another peer. Reputation is based on what other peers say about a particular peer. Trust and reputation are developed in a different manner, yet [11, 12 and 13] they share the following common characteristics:

Context Specific. Trust and reputation depend on some form of context. A person X may trust a person Z as his dentist. However, person X may not trust the same person Z as his plumber. The level of trustworthiness of the same person differs in two different contexts.

Multi-faceted. Trust and reputation are not necessarily developed on the basis of a single factor. Different attributes build or break trust or reputation. For example, the reputation of a brand producing cars may be a combination of safety, aesthetics, innovation and consumption.

Dynamic. The more experience, the more trust and reputation tend to change. Trust and reputation may increase as well decrease by time.

Content retrieval is one of the major activities in a p2p network. Such content retrieval processes involves two phases: a *content search* phase and *content download* phase [10]. In *Content search* a peer sends a query with the requested content to the network. This hops from one node to another until its assigned TTL (time-to-live) is 0. The TTL decreases from one hop to another. The nodes having the requested content will reply to the initiating. Then, the *content download* process is started. The requesting peer connects to one of the selected peers as a result of *content search*. The content is then downloaded in the *content download* phase.

Malicious nodes may intervene at any step of this whole process. In a p2p network, there has to be a level of co-operation between peers. Bad behavior of corrupt peers hinders such co-operation. A malicious peer may choose not to forward the query to the rest of the network. The content of the query may be changes before forwarded to the rest of the peers. The requesting peer may be informed that the required content is available and then receives completely different content such as malware. A proper trust and reputation system can help decrease such undesirable actions. A misbehaving peer will eventually get a bad reputation. A node having a bad reputation will not be chosen by the other peers for content download.

Even though reputation seems to be a nice solution, there are still problems, which have to be taken in consideration when creating such a system [14].

Pseudonyms. To ensure anonymity, certain p2p networks use pseudonyms to identify nodes in a network. A new identity can be simply created by creating a new pseudonym. Therefore, peers with bad reputation can easily rejoin the network.

Pseudospoofing. If it is possible to create multiple pseudonyms for the same peer, then it is possible for a bad node to create other peer identities in order to get a high reputation within the network.

Shilling. This is similar to pseudospoofing in that different real IP addresses are used to create the different identities on the network.

Cold-start. Peers having high reputation are chosen for content download. A user joining the network will start off with no or very low reputation. This will make it very difficult for such peer to be chosen for content download.

Load problems. Since peers with high-level reputation will be more chosen than those having low-level reputation, high-level reputation peers will have a higher load. A proper load-balancing method should be implemented to avoid over-loading such peers.

Different papers discuss different models of trust and reputation systems. A reputation system based on *DCRC (Debit-Credit Reputation Computation)* and *CORC (Credit only Reputation Computation) is* discussed in [a]. The peer can select not to have his activity tracked and its reputation will always be 0. The DCRC method gives the peer credit for making content available on the network and debits the peer for downloading. The CORC gives only credit when content is made available. Under both schemas, extra credits are given to peers for query processing, forwarding and staying online. The major drawback of this system is that it uses a central *RCA (Reputation Computation Agent)* to credit reputation points. This reduces the level of decentralization in the network. Also, in CORC peers can get a high a reputation by transferring content to and from each other.

Xrep is another reputation system that is discussed in [a]. Together with peer reputation, this system takes also in consideration file reputation. Earlier in this paper, we discussed the two-phase process in the content retrieval process: a *content search* phase and *content download* phase. Xrep extends this to a 5-phase system. In the *first phase*, the requesting peer sends a request for content. Nodes, which can satisfy such request, notify the requesting and also send a digest of the file that matched the keywords included in the query. In the *second phase*, the peer for downloading is selected. Reputation is calculated with a broadcast of the chosen peer and the file to the entire network. Nodes on the network reply with their view. The *third phase* involves the requesting peer analyzing the votes. To avoid shilling, each voting peer provides the IP address and is contacted. In the *fourth phase*, the winning peer is contacted and confirms that the data can be exported. The *fifth* and final stage consists of actually downloading the content and checking

the integrity of the file. The opinion of the requesting peer on the content of the file and the serving peer is updated accordingly.
Anonymity is the main problem here. Voting peers are providing their IP address. This exposes their identity and malicious peers can attack peers reporting a bad value. However, such system does not suffer from cold-start problems. New users can participate in the network once they join if they offer content that is highly requested. TrustMe is a trust management system which is also discussed in [a]. Assigning trust anonymously is the main aim of this. It uses THA's (Trust Holding Agents). When a new peer joins the network, a THA is assigned to it. The THA will be unknown to all of the peers connected to the network, even to the newly joined peer. The THA keeps track of all the activity of such node on the network and adjusts the trust values accordingly. Trust information on a peer can be gathered by broadcasting a trust query. The THA assigned to that peer provides the necessary information.

A proposed trust model is based on five main components: the users (U), Trust (T), Experience (E), sharing files (F) and heterogeneity level (L). As explained in [d] definitions for these components are:

- $U=\{u_1,u_2,...,u_n\}$ the set of peers in the peer-to-peer network, where n is the number of peers
- T is an attribute of a peer, which denotes the trust value of a peer and belongs to the interval [0,1]
- E is a value measured by the contribution of a peer in the network.
- $F=\{F_1,F_2,...,F_m\}$ is the set of sharing files by a peer, where m is the number of files
- $L = \{L_1, L_2, ..., L_k\}$ is the set of heterogeneity level in the network.

Different heterogeneity levels are assigned to the peers. Peers with the same heterogeneity are grouped together, thus creating different levels of peers with different heterogeneity. This can be seen in Figure 1.

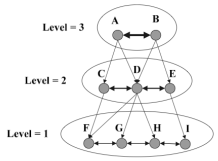

Figure 2. The hierarchical structure of peers

Peers at level one will have the less power. Peers at the top level, such as Peer A, can access the files of the peers at the bottom levels, such as B, E, I and so on. Such a framework offers solutions to various problems. For example peers may secretly team up in order to increase the reputation of each other, thus having other peers trusting in them. Due to the factor of experience value, even though peers can conspire to enhance their trust value, but the sum of the experience value of these peers cannot be increased [d]. Also such a model tries to avoid free-riders. The only way how one can access more files is by going up in the heterogeneity levels. If a peer does not contribute to the network, then it will not go on higher levels, does limiting the variety of files available to download.
[c] proposes a Bayesian network-based trust model. The basis of such trust model is that peers share the experiences that they have on the network with each other. Peers will take decisions depending on the experience of other peers on the network. Two types of trust are developed, the trust in a peer's capability to provide content and reliability of on peers in making proper recommendations. After a requesting peer requests for content, it compiles a list of peers which can provide such content. If requesting peer has a good experience with the top peer, then the process is continued. However, if not, it can ask other peers to make recommendations of such peer. What the peer does after that is its decision, depending on the level of cautiosness that it applies. If the content is downloaded, the requesting peer updates the relevant evauations, depending on the interation it had. Peers which refer other peers are called "referees". If the requesting peer performs the transaction based on the referees recommendation, then the trust evaluation of such referees is also updated. Different peers may value trust on different combination of factors. For example certain peers may put more importance to download speed while others to the file quality. In [a] a Bayesian network with three factors is created. Trust is based upon Download Speed, File Quality and File Type. According to a Bayesian network, a peer can set the trust depending on these aspects. The method will save peers effort in building different trusts separately, or developing new trust when conditions change.
A group is described as a set of peers that are governed under a set of rules that describe minimal conditions of a group [b]. In this paper we see a model for peer-to-peer access control with the aim of having more secure networks. The computation of Group-to-Group trust is built on PPT (peer-to-peer trust). A Trust Matrix, first maps the PPT. The level of trust from 0 to 1 between each and every peer is recorded. Values which are 0 mean that a transaction between peers failed. Null values are present if no transaction take place between peers. Also, the trust is unidirectional. This means that PPT (P2, P5) ≠ PPT (P5, P2) Peers are then split in groups in order to come up with the GGT (group-to-group trust). Peers in a group only express their trust with the peers within that group. GGT is also unidirectional as PPT. This paper is proposing that by using group trust for access control may reduce some of the problems which we face in peer-to-peer computing. Without access control, if a peer is being overloaded by requests from another peer, then these requests have to processed. However, if the requesting peer comes from a group with low trust and reputation, then such requests would be rejected.

III. EXPERIMENTS AND RESULTS

In [d] an exercise was done in a peer-to-peer network with 1,000 participants. Half of the nodes on the network are good peers while the other half are malicious peers. First, 10,000 transactions are randomly performed between peers. Then peers are set to levels depending on their trust and experience values. After that, transactions are made by each peer until 60,000 transactions in total are reached. The results show that at the beginning, malicious peers are 'kind' in order to incease their reputation. The transaction successful rate decreases then. This happens when malicious peers start harming the network. However, the bad peers start uncovering their identity. Thus, the transaction successful rate starts going up again.

Figure

An experiment to prove Bayesian network of trust is also done in [c]. Every peer has an interest based on five elements: *music, movie, image, document* and *software.* Each element holds a value which indicates the strength of the peer's interests in the particular file type. Every peer keep a list of the other peers which had interaction with and its trust value on these peers. Another list keeps a track of file providers and their corresponding Bayesian networks. In the setup there are 10 file providers and 40 requesting peers. There are a total of 10,000 transactions and peers will exchange their Bayesian networks after every 5 transactions. First this test is run with Bayesian networks and then it is run without.

As can be seen in **Figure 1,** a system with Bayesian networks perform better than those without. In a second experiment, another test is performed. However, we compare a system sharing trust with and without Bayesian networks and another sharing both trust and reputation with and without Bayesian networks.

Figure 2 shows that peers who share information about each other do better than those systems which do not. In some sense, a peer's Bayesian network can be viewed as the model of a specified file provider from the peer's personal perspective [c]. In these experiments, the Bayesian network is quite simple. It is only based on 3 factors: *download speed, file quality* and file type. A more complex network would be required in order to be used in real-world file-sharing situations. In fact, such a Bayesian network is ideal in a small peer-to-peer network in which a lot of peers are coupled interaction. In order for this to be successful in a larger network scenario, the small-world phenomenon has to happen. Small-world simply means that even on a peers are inclined to get files from other peers from a small sub-community. This small sub-community often consists of peers that have similar preferences and viewpoints [c].

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. ans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.