Attacks Against Peer-to-peer Networks and Countermeasures

Lin Wang
Helsinki University of Technology
lwang@cc.hut.fi

Abstract

Peer-to-peer (P2P) networks have many distinct aspects that are different from traditional client-server networks. The most significant point is that each peer acts as both server and client roles in P2P network. In other words, there is no central server that used for storing the files and providing download. All nodes download files directly from other peers. P2P networks contain ad hoc, decentralized structures and autonomy peers. Each peer can randomly leave or join in the network and the network topology is changed every now and then. These characteristics of the P2P network make it vulnerable. Thus, the security issues of P2P networks is a serious topic that should be considered carefully. In this paper, we study the general P2P system structures, attacks that may occur in the different P2P topologies and the potential countermeasures against those attacks.

KEYWORDS: peer-to-peer, overlay network, distributed hash table, DoS attack, poisoning attack, anonymous overlay network, Sybil attack, eclipse attack

1 Introduction

A P2P networks are dynamic and distributed networks. Normally, the main difference between the concepts of P2P network and traditional server-client network is that the file downloading is not provided by a central server. Nowadays, P2P technology is widely used for files sharing, instance message communication and distributed computing.

According to the network topology structures, current P2P network can be divided into four categories: Centralized P2P (Napster), Pure P2P (Gnutella 0.4, Freenet), Hybrid P2P (Gnutella 0.6, JXTA) and DHT (Distributed Hash Table)-based P2P (Chord) [1].

The security issues are inherent features accompanying with P2P systems. It is highly suscepted to various forms of malicious attacks. It can not only be attacked from the malicious nodes outside the P2P network but also vulnerable to its own peers. The thousands to millions of anonymous peers provide an ideal attack environment for attackers. In addition, the popularity of P2P also leads different kinds of security issues.

Similar to traditional Internet, P2P networks are vulnerable to many general attacks, such as DoS attack, DDoS attack, poisoning attack. It can also be the victim of Sybil attacks and eclipse attacks. Attacks could happen on both application layer and network layer. In this paper, we pay more attention to the different network layer attacks and countermeasures based on the structures of P2P networks.

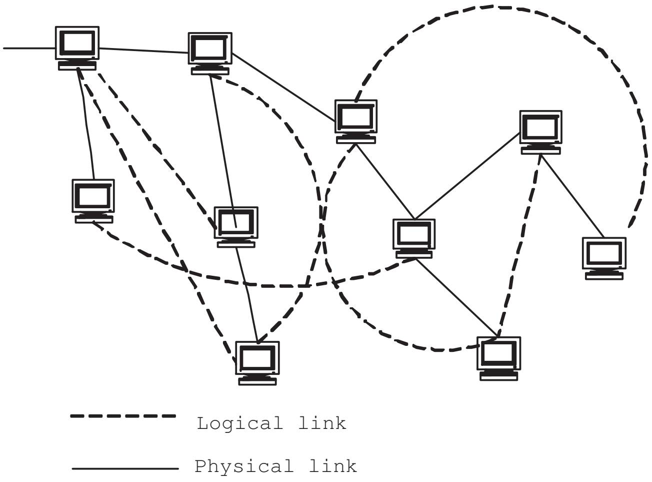

Figure 1: Overlay network

The rest of paper is organized as follows. Section2 introduces the general background knowledge about the structures, technologies and main working schemes of the above principle P2P techniques and analyzes their vulnerabilities and the potential security issues. Subsequently, Section3 focus on the details of main attacks in P2P networks and solutions to prevent or decrease the damages. Finally, section4 makes a conclusion about the the overall article and the situation of current P2P networks security.

2 Structures of P2P

In this section, we introduce some background knowledge about several common P2P systems. P2P system is a kind of overlay network. Figure 1 shows the concept of overlay network. In despite of the physical connections, virtual overlay network is built based on the virtual connections of nodes.

P2P systems can be classified to unstructured P2P and structured P2P. In unstructured P2P system, files can be stored in any peer, that is, the file storage has no certain structure. In contrast, structured P2P system maintains a link between file contents and IP addresses of the peers using Distributed Hash Table (DHT) [5]. Therefore, the whole P2P system holds a certain structure. Unstructured P2P system includes Centralized P2P, pure P2P and Hybrid P2P. Normally, structured P2P means DHT-base P2P.

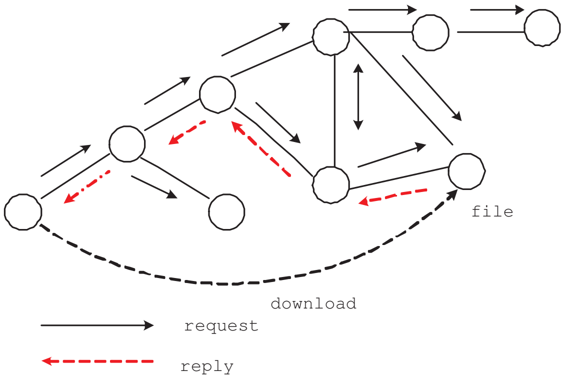

Figure 2: Flooding query of pure P2P

2.1 Centralized P2P

Centralized P2P utilize a central server to provide only the index of the distributed files to the peers. Therefore, in centralized P2P, each peer uploads the file index to the central server. When a peer wants to download a file, it sends query to the index server to request for the file. Then the central server will find an ideal available peer that contains the file and send the reply with the peer's address. Although the query approach of centralized P2P relies on a central server, but after getting the content destination, the query peer downloads files directly from the destination peer by HTTP.

2.2 Pure P2P

Contrary to the centralized P2P, pure P2P has no central serverat all. As shown in Figure 2, A peer requests the desired content by broadcasting the queries to all its neighbors. Then its neighbors continue flooding the requests to the further nodes until the Time to Live (TTL) is exceed. When the peer which holds the needed file gets the request, it responses the query then a download session will be established between initial peer and destination peer. The drawback of pure P2P is that it generate a huge amount of signaling traffic by flooding. And also the routing mechanism is based on best effort, therefore, a query node may not get the reply even the destination is available in the network.

2.3 Hybrid P2P

Hybrid P2P combines the features of centralized P2P and pure P2P. It defines several superpeers. Each super peer acts as a server to a small portion of network. Moreover, each superpeer stores a list of index files information that are available to the peers that it manages. Hybrid P2P uses the similar query mechanism as centralized P2P. In addition, the superpeers only manages a small range of sub-peers and stores index information for its sub-peers. This feature makes it answer the queries quickly. Compare with the above two schemes, hybrid P2P reduces the signalling traffic and saves the badwidth.

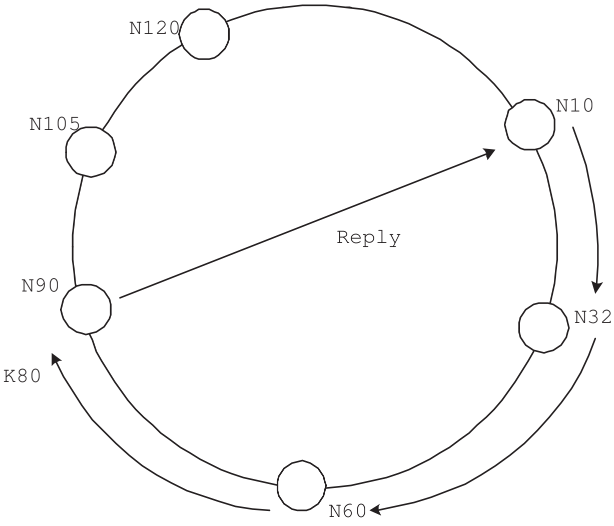

Figure 3: DHT circular address space and basic look up

2.4 DHT-Based P2P

DHT-based P2P is a structured P2P system. In DHT-based network, nodes are organized in a structured overlay. The Figure 3 illustrates a linear circular address space and the basic look up of Chord. In the picture, each DHT node maintains a range of address space and the address spaces can be consider as a circle. Each file in DHT system is assigned an unique ID in a certain address space. The ID consists of two parts: key and value. Key is the hash value of file name, value is the hash value of file content. Each node in DHT manages a range of key values that start from itself to the successor node that contains the smaller key value. For example, the node 10 manages the key values in the interval 10 to 31.

By using the DHT algorithm, the peers can map the keys to nodes easily. The basic look up scheme is shown on Figure 3: a request peer looks for key 80 from the node 10. Each node in the picture maintains a range of view of the system. When it (N10) gets a destination ID (K80) that is not responded by itself, as per the identifier sequence of the DHT address space, it forwards the message to the nodes (N32) that close to the destination. If the next node (N60) is not the destination, it will forward the message to the next node (N90) that closer to the destination [5]. Although different DHT-base P2P system may use different routing scheme to look for the next hop, but the only rule they have to obey is that message are forwarded to the peers whose identifiers numeral close to the destination.

3 Attacks and Countermeasures

A P2P network has an unstable structure because it allows the unreliable nodes to frequently access or quit the system. Moreover, the decentralized property makes the traditional "smart server" security mechanism, such as, VPN can not work in P2P environment [11]. Therefore, the security issues in P2P networks are important and challenging. This section specifies the usual malicious activities and countermeasures in P2P network.

3.1 Routing Attacks in DHT Based P2P

In DHT based P2P network, each node maintains a routing table and the routing table guarantees the look up and mapping of the keys. A malicious node which serves as an active part of P2P network can perform some abnormal behaviors. An ordinary attack is that the attacker forwards the look up request to an incorrect node.

According to the property of the DHT, this kind of attacks are easy to detect. We make the query node be aware of that the look up should be get closer and closer to the key identifier. If the query node found the look up does not follow this rule, it should find the last correct node and ask for another route [3].

3.2 DoS and DDoS Attacks

Denial-of-service(DoS) is a common attack but difficult to be prevented both in traditional internet and p2p networks. In DoS attack, attackers utilize reasonable service requests to exhaust the resources of a target host. Therefore, the victim host can not provide any service to other legal intended users. Distribute DoS (DDoS) is an attack developed from DoS. The purposes and the concepts of DDoS are same as DoS. However, the specific method of DDoS is different. DDoS attacker exploits considerable amount of distributed hosts to launch the attacks to the target. In other words, it is a larger scale distributed DoS.

P2P networks are composed by large number and anonymous concurrently running hosts. Thus, one or more malicious nodes in the network can easily perform DoS or DDoS attacks to a victim server or client. DoS or DDoS attacks may have different levels of influence base on various P2P network topologies. The P2P networks with central server, such as, Napster get more serious result after encountering the DoS (DDoS) attacks. Because the centralized P2P has single point failure, thus, the failure of the central server directly leads the whole network down. In contrast, DoS causes less damage in the pure P2P and hybrid P2P. The failure nodes only affect a part of network. Anyway, DoS makes the victim node can not provide the required service to desired clients any more and the network performance decreases. That is, P2P networks are vulnerable to DoS (DDoS) attacks, for example, TCP Syn flooding and query flooding.

3.2.1 TCP Syn Flooding Attack

Classic TCP Syn flooding attack can also be applied in P2P network as that in traditional network. It may happen in the stage when the attacker is downloading files from the victim host. Because the downloading process in P2P networks are accomplished by using TCP connection, the vulnerability of three way handshakes of TCP make the peers could be a target of DoS (DDoS) attack. The attackers firstly use a forged IP address to send a SYN request to the target host. After getting the SYN request, the victim host responses with a SYN-ACK message. However, the IP address of the attacker is fake, so the third handshake could not be accomplished. In other word, no ACK message will be sent to the victim node forever. The attackers continuously send a great amount of SYN requests to victim. Finally, the victim host exhausts the resources and can not perform service to the legal peers.

The TCP Syn flooding attack is also known as half-opened TCP Syn attack, because all the TCP requests from attacker remain half open status. Syn cookie is a good method to against this attack. When a server gets the TCP Syn packet, it hashes and encrypts some security values, such as client IP address and port number to get a initial sequence value. It does not drop the new connections even if the syn queue of the server is full in DoS attack. In contrast, when the server get another Syn request, it also replies SYN + ACK. In former processes, the server does not allocate any data section. When the server receives an ACK, it verify the ACK by hashing then comparing with the initial sequence value. If the value is the same, the server allocates data section to handle TCP connection. The limitation of syn cookies is that it only can be supported by Linux and Solaris operating system.

3.2.2 Query Flooding Attack

The query flooding attack can also happen in application layer in some pure P2P systems, such as Gnudella. In order to obtain the desired files, the query node has to broadcast the queries to all its neighbors. A malicious node will constantly generate as much queries as possible to flood the network. If the DDoS query flooding attacks happen in centralized P2P, for example, Napster. The harm will be bigger. In Napster network, if a lot of malicious peers continuous send queries to the index server, the great heavy traffic can overwhelmed the server. That causes the file query processes can not be finished. Consequently, the downloading session can not be established and the whole P2P network does not work. Actually, such kind of query flooding attack is easy to control. Usually, it can not break the whole network but only reduce the network performance in Gnutella. Because each node in Gnutella knows a maximum number of queries of a good node. Therefore, a node can accept at most the maximum queries from a request peer. After getting the maximum number of queries from a request node, it just drops the rest requests from that incoming link [6]. This mechanism can effectively decreases the harm of the query flooding attack but can not totally avoid it.

We imagin another kind of distributed query flooding that may happen in pure P2P networks. A set of distributed attackers intend to attack a victim node. They use the victim's IP address as source address and send queries to ask for some popular files. The destination nodes which contain the files could be another set of distributed peers. Each destination peer sends its reply to the same fake source address. That means, the victim peer has to accept a huge amount of replies from different servers and may overwhelmed by those traffic. The countermeasure to non-distributed query flooding attack can not solve this problem successfully. Because the number of queries from each attack is less than the maximum value, the victim does not do any reaction to deny the incoming traffic. But the total amount queries from attackers are a disaster to the victim node.

3.3 Poisoning Attacks

Poisoning attacks can happen in the P2P networks. Attackers use false information, for example, false file indexes, false IP addresses, or false routing tables, to break the integrity of P2P systems. We introduce two kinds of poisoning attacks, index poisoning attack [4] and routing table poisoning attack [4].

3.3.1 Index Poisoning Attack

Some P2P networks employ central server (e.g. Napster) or distributed server (e.g. Skype) to store the index of files. Index poisoning means the attacker adds considerable bogus index information into the index server. The most important thing is all bogus information point the target address of the popular files to one target victim host. When other peers attempt to download those popular files, they get bogus information from the index server. Then those peers establish the TCP connection to the victim node [4]. In this case, the other peers can not get services from the victim node because the fooled peers have occupied the allowed connections. In addition, the victim node can easily be exhausted by the TCP connections between the fooled requesters.

The countermeasure of the index poisoning attack is difficult to find in the current stage. Because a full TCP connection is completed between the victim node and each requester. The TCP Syn cookies can not solve this problem. It is also very difficult for servers to recognize their potential clients in TCP connection stage.

3.3.2 Routing Table Poisoning Attack

Routing table poisoning are practiced in DHT-based P2P networks, such as, Chord and CAN. The idea of routing table poisoning is the malicious attackers add the victims' IP addresses into the routing tables of a set of peers as their neighbors. The poisoning process may be managed by sending bogus messages that clarify the existence of the victim IP address to different peers from the attackers. Then those peers include the victim node as their neighbor [4]. In forwarding process, those peers may choose the victim host as a relay then send the packets to the victim. Consider if we have millions of such fooled peers, the victim can be easily exhausted and crashed. The routing table poisoning is not a deadly attack for P2P systems. Because many P2P system hosts can update the routing table automatically. When the peers can not get responses from their neighbors, they will delete that bogus IP addresses from the routing tables.

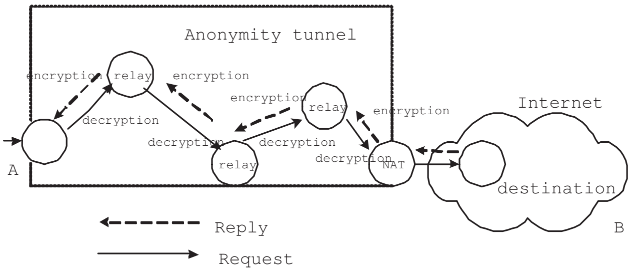

Figure 4: Network overview of Tarzan

3.4 Tracking Attack and Anonymous Overlay Network

Attackers can track and find out the IP addresses of clients or servers then generate attacks to them. One method which can well contraposes such kinds of IP addresses attacks, for example DoS attack, and provides location privacy is known as anonymous overlay network. The main idea of the anonymous overlay network is to protect the network by concealing the network addresses of the participants from the malicious nodes. Recently, there are three categories of anonymous: "server anonymity" or "responser anonymity" protects the security of servers; "client anonymity" or "sender anonymity" protects the security of clients; mutual anonymity can ensure both the client and server anonymity at the same time [2]. In the following chapter, We introduce a type of anonymous mechanisms which is known as Tarzan [10].

Tarzan provides "either clients or servers anonymity". In addition, it can achieve mutual anonymity in case both the sender and recipient are involved in Tarzan system [10].

Figure 4 shows the network structure and working principle of Tarzan. Node A is a Tarzan node running a client application [10], in another words, it is the node requires anonymity. It selects a group of available peers to compose the transport path. All these peers called "relay" [10] are selected randomly for security reason. After the nodes selection, the symmetric keys of each relay are distributed. Then, node A encrypts packets several times using the symmetric keys of the relays. All these nodes compose a forwarding path that is considered as a tunnel. Thus, the encrypted packets start from node A and go through the relays. Each relay decrypts one layer of encryption and forwards the packets to the next hop, until the packets reach the last relay which runs a network address translator (NAT). By this way, each relay can only know its previous and succeed nodes but have no idea about the source node. The last node decrypts the last layer encryption and obtains the destination address of the packets. Finally, the packets are sent to the public destination address by the NAT node. The reply path is a reverse tunnel of sending path. The final node encrypts the message using its symmetric key and forwards it to the penultimate relay. Then the relays encrypt and send the message one by one until it gets to Node A. Node A decrypts the layers of encryption and gets the message. By using Tarzan, the anonymity of the nodes is well protected. However, a notable disadvantage of Tarzan is that it utilize best effort routing principle.

There is another proposal about anonymous overlay network called Anonymous Server Overlay Network (SASON) [2]. Compare with Tarzan, SASON uses different structure and concept to manage the server anonymity. Moreover, It provides reliable routing by using Distribute Hash Table (DHT). The SASON framework can be considered as being composed by DHT routing network and anonymous sender routing network which hides a server from its clients and also the malicious attackers [2]. We do not mention more about that in this article.

3.5 Sybil Attack

P2P networks are virtual overlay networks built on an underlay network. That means each entity in underlay network has an corresponding identity in overlay networks. An identity should be unique and can form a one-one mapping pair to actual entry. Most P2P networks use the virtual addressing scheme based on the logical identifiers to manage and organize the network [11]. However, if the relation of one-one mapping of entity to identity is destroyed by malicious peer, in other word, a malicious entity acts as a number of multiple identities. The entity can control a significant part of networks. Such attack is defined as Sybil attack [7]. Sybil attack can occur in all the networks that require the entity and identity mapping, such as, P2P networks, ad-hoc and sensor networks [11].

The primary compromise of Sybil attack in P2P network is destroying the redundancy. Many P2P systems rely on replicating the computational or storage tasks and segmenting tasks among several remote peers to protect the integrity and privacy [7]. In this situation, with the Sybil attack, the system may finally select only one node that act multiple identities. The redundantly performance is defeated by Sybil attack

A way to resolve the Sybil attack in P2P is establishing a trusted certificate authority to make distinct entity has distinct identities. There are several existing methods, for example, the explicit certification agency, VeriSign; CFS cooperative storage system that identify node by a hash of network address; EMBASSY, utilizes cryptographic key in hardware peers to identification [7].

Currently, there is no perfect idea to resolve the Sybil attack. We briefly introduce an "identity registration procedure"called "Self-registration" to against Sybil attack [11].

In order to calculate their identifiers, first, node hashes its IP address and port as ID. Then it registers the ID at already registered nodes. The registration process is accomplished and the node can join the P2P networks and other nodes have ability to verify the ID and distinguish whether it is a fake ID.

3.6 Eclipse Attack

Eclipse attack is a general attack in overlay networks [8]. In eclipse attack, an attacker controls a large part of the neighbors of a good node. In this situation, the union of malicious nodes work together to fool the good node by writting their addresses into the neighbor table of a good node [8]. By using eclipse attack, attacker can control the significant part of overlay network. In addition, large scale malicious nodes can eclipse more good nodes to control the entire overlay network. The overlay nodes can not forward message correctly and the whole network cannot be managed. A Sybil attack can be considered as a specific Eclipse attack, if the attacker generates great amount of identifications to act as neighbors of a good node.

In Eclipse attack, the indegree of an attacker must be higher than the average level of the indegree of nodes in a P2P network. Here, indegree means the number of direct routes coming into a node and outdegree means the number of direct routes going out of a node. Good node can choose the neighbors that the indegrees are below the threshold. But malicious nodes can also make a poisoning attack by decrease the indegree of good node. Hence, the good node will never choose a neighbor. So, the idea is to bound both indegree and outdegree of attacker nodes.

We present a method that introduced in [9] to prevent eclipse attack. First, we can apply the countermeasure to the Sybil attack. That process assure there is no possibility of Eclipse attack which based on a Sybil attack. Then we concentrate on how to deal with the indegree and outdegree of the attacker nodes. Each node in P2P networks maintains a list of its neighbors. We make the node periodic query the neighbor lists of its neighbor peers. If the the items on the replied neighbor list are greater than the indegree bound, or that node is not on his neighbors' list or the size of returned neighbor is greater than the outdegree bound [9].

4 Conclusion

This paper presents some common attacks and countermeasures in current P2P networks. In this paper, we discussed the different categories of P2P systems and analyzed some basic attacks. Some of them are usual attacks in the Internet, such as, poisoning attacks and flooding attacks. Others are particular attacks against overlay networks, such as Sybil attacks and eclipse attacks. The defenses of the attacks are also clarified.

Some attacks are easy to generate, such as, TCP Syn attack, query flooding attack, poisoning attack and sybil attack. Attackers don not need to spend much energy but can get effective result. TCP Syn attack can be solved by utilizing syn cookies. Most of the P2P system have such mechanism to against query flooding but distributed query flooding is still a headache topic to investigate in the future. Both structured and unstructured P2P prototypes are highly vulnerable to index poisoning attack. It is easy to generated by the attackers, because what an attacker needs to do is just inserting bogus index records. But it is very difficult for victim peers to recognize and defend. Compare with index poisoning, the routing table poisoning can not make enormous harm to current P2P topologies. Because most of the P2P system can update the routing table and delete the bogus neighbors automatically.

We can propose a hardware based security method to against poisoning attacks. Trusted Platform Module (TPM) chip [12] can be utilized to provide remote attestation between P2P network nodes. Remote attestation allows a third authorized party to detect the changes of the software on a computer [12]. This mechanism avoids the attackers to tamper the file indexes or routing tables. Certainly, the hadrware based remote attestation requires all the nodes in the P2P network to support the TPM chip. Today, the IBM laptops already include the TPM chip.

Sybil attack and eclipse attack are really two big threats to P2P network, in particular of DHT-based overlay network. Because the attackers can control a significant part of the network. In addition, there is no mature and complete mechanism to against them.

Anonymous overlay network can hide the IP addresses of the peers. It provides location privacy and can effectively defend some IP address based attacks or decreases the opportunities of such attacks. But such P2P network is expensive to implement because of the cryptography.

The current P2P systems are still very weak in security. Moreover, the proposed methods are not ideal enough and could reduce the performance at the same time of enhancing the security. The development of the security mechanisms in P2P networks are still an important investigate topic in nowadays. Only when the ideal defense systems are built, the P2P system could have a better development.

References

[1] Jörg Eberspächer and Rudiger Schollmeier. First and Second Generation of Peer-to-Peer systems. Springer-Verlag Berlin Heidelberg, 2005.

[2] Henry Tsai and Aaron Harwood. A Scalable Anonymous Server Overlay Network. *Proceedings of the 20th International Conference on Advanced Information Networking and Applications*, 2006.

[3] Emil Sit and Robert Morris Security Considerations for Peer-to-Peer Distributed Hash Tables. *W*orkshop on Peer-to-Peer Systems,March 2002

[4] Naoum Naoumov and Keith Ross. Exploiting P2P Systems for DDoS Attacks. *Proceeding of the First International Conference on Scalable Information Systems*, Hong Kong, June-2006.

[5] Klaus Wehrle and Simon Rieche. Distributed Hash Tables. *Springer-Verlag Berlin Heidelberg*, 2005.

[6] Neil Daswani and Hector Garcia-molina Query-Flood DoS Attacks in Gnutella. *I*nACM CCS, 2002

[7] J.Douceur. The Sybil Attack. *Proceedings of the First International Workshop on Peer-to-peer Systems*. Springer, March 2002.

[8] Atul Singh, Tsuen-Wan "Johnny" Ngan, Peter Druschel and Dan S. Wallach. Eclipse Attacks on Overlay networks: Threats and Defenses. *N*gan et al., Infocom 2006.

[9] Atul Singh, Miguel Castro, Peter Druschel and Antony Rowstron. Defending Against Eclipse Attacks on Overlay Networks *ACM. In Proc. of the 11th European SIGOPS Workshop*, Leuven, Belgium, September 2004.

[10] Micheal J. Freedman and Robert Morris. Tarzan: A Peer-to-Peer Anonymizing Network Layer. *ACM Conference on Computer and Communication Security*, Washington, 2002.

[11] Jochen Dinger and Hannes Hartenstein. Defending the Sybil Attack in P2P Networks: Taxonomy, Challenges, and a Proposal for self-Registration. *IEEE, Proceedings of the First international Conference on Availability, Reliability and Security*, 2002.

[12] Platform Module, Wikipedia, 2006
http://en.wikipedia.org/wiki/Trusted_Platform_Module