Modeling Leechers Attack in BitTorrent

Lin Ye, Hongli Zhang
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China, 150001
Email: hityelin@gmail.com, zhanghongli@hit.edu.cn

Xiaojiang Du
Dept. of Computer and Information Sciences
Temple University
Philadelphia, PA, USA
Email: dxj@ieee.org

***Abstract*—As one of the most widely-used Peer-to-Peer applications, BitTorrent system has gained a lot of successes in last decade. BitTorrent has been extensively studied in the literature. However, few research studies vulnerabilities of BitTorrent, and the abuses of BitTorrent have already impeded the wide adoption of the system. In this paper, we study a common attack on BitTorrent - the Leechers attack, which tries to take over valid connections of legitimate users without paying for them. It is harmful to a swarm if certain percentage of users perform the selfish behavior. We present an analytical model for the Leechers attack and we use the model to estimate the distribution of valid connections. Our model discovers the important factors that determine how harmful the attack is, and it can quantitatively predict the duration of downloading process under the attack. The real-world experimental results match well with our model, which demonstrates that the model is correct and useful.**

***Index Terms*—Leechers attack, BitTorrent, Model**

I. Introduction

Peer-to-Peer (P2P) technologies have been widely used in last decade. P2P provides an efficient way to share resources over the Internet. Particularly, BitTorrent (BT) [1] attracts more and more users due to its good scalability and fast downloading. BT has been widely used in other fields [2], [3] as well. According to an early study [4], there are millions of users in BT network. Unfortunately, with the increasing popularity of BT, the risks of malicious behaviors exploiting the potential vulnerabilities are also on the rise. Many efforts have been made to improve BT system by introducing and enhancing incentive mechanisms. However, BT system is still not robust enough and many users are suffering from the performance degradation and even attacks [5]. Furthermore, BT system has been exploited as a tool to launch DDoS attacks [6]. Several measurement and simulation results have shown the impact of malicious behaviors on BT. However, it is unclear what factors determine the BT system efficiency and how they influence the system.

Different from previous work, we theoretically model one of the most common attacks on BT system - the Leechers attack [7] [8] in this paper. We also estimate its impact on BT. In the Leechers attack, attackers can obtain resources at a faster speed without uploading any data. This obviously damages the fairness among users and discourages them to share resources. Leechers attack exploits BT in two aspects: first, the vulnerability of authentication allows an attacker to easily generate thousands of forged identities using only a small number of machines. Attackers can increase, decrease or change those identities dynamically. Second, the random selection mechanism in optimistic unchoking algorithm allows a peer to obtain connections from others even if it never uploads blocks. A good model of the Leechers attack can discover important factors that determines the degradation level, and then can predict the duration of downloading process under an attack.

The rest of this paper is organized as follows. Section II describes BT system and summarizes the related work. Section III introduces the analytical model of Leechers attack. Section IV presents the real-world experiments and discusses possible countermeasures, followed by the conclusion in Section V.

II. Background

A. BitTorrent System

BitTorrent system consists of four parts, including torrent index, peer index, seeds, and leechers. A torrent is the meta data that stores description information of a file. A torrent is important for the users because it contains necessary information to bootstrap users into a collection of peers (also called a swarm). A torrent index is a set of ongoing torrents that are collectively organized in the form of torrent websites like famous Mininova [9] and BTChina [10]. These websites allow users to upload their torrents and provide tracker services. A peer index is a set of peers that participate in the distribution of a specific file. The basic function of a peer index is to track the status of peers that are currently active, and act as a rendezvous point for all peers. There are three index schemes in BT, i.e., tracker, Distributed Hash Tables (DHTs) [11], and Peer EXchange (PEX) [12]. In the tracker scheme, each peer registers itself to a server called tracker, and each peer may also obtain a random subset of other peers (IP addresses and Port #). DHT is used to support distributed index maintenance. In DHT, a client (called a DHT node) can query “infohash” to obtain a similar result of the other peers where “infohash” is a unique fingerprint of a .torrent file. PEX is an alternative way that allows the peers to exchange the information of their active neighbours with each other.

Depending on their download states, peers are classified into two types: seeds and leechers. A seed is a peer that has already downloaded a file and is willing to serve the other peers even though it does not need any more contents. A leecher is a peer that has downloaded part of the file. A leecher provides part of

978-1-4799-3512-3/14/$31.00 ©2014 IEEE

the file (that it downloaded) to the other peers and meanwhile it downloads the rest of the file from the other peers.

To encourage collaboration among peers, Cohen [13] proposed a “Tit-for-Tat” incentive mechanism to prevent selfish “free-riding” behaviors. In general, each peer has many simultaneous connections with others, but it can upload part of them (default number is four). A peer gives preference to higher bandwidth peers that have uploaded to it before. In addition, a peer reallocates an upload slot to a random peer every 30 seconds, which is also called optimistic unchoking. Optimistic unchoking has two advantages. First, it is favorable to seek for the peers with higher download rates for better performance; Second, it allows new peers to obtain upload slots though they have not uploaded any content to others yet.

B. Related Work

The vulnerabilities of BT protocol have been studied in the literature. In order to defend against selfish peers, the “Tit-for-Tat” mechanism is used to encourage the high-bandwidth peers by evaluating the bandwidth of different peers [13]. However, BT protocol is still not robust enough. Study [7] proved that users could benefit from three selfish downloading behaviors: only downloading from the seeds, only downloading from the fast peers and downloading by cheating on fake upload blocks. Work [14] found that BT cannot be immune to “Free-riding” behavior, and designed a new incentive mechanism to encourage the cooperation among peers. Work [15] also implemented the client “BitThief” that downloads from others without any upload. Study [8] analyzed the vulnerabilities of optimistic unchoking algorithm, and pointed out that as long as the number of connections to other peers is large enough (Large View), optimistic unchoking can “help” malicious peers to obtain the data without any cost. Work [16] maximized download bandwidth at the cost of minimal upload bandwidth by optimizing peers selection, tuning parameters, et al. Their effectiveness indicates that “Tit-for-Tat” mechanism is not robust enough.

Meanwhile, the measurement results [5] show that many peers are attacking or polluting BT system, such as Fake-Block attack and Uncooperative-Peer attacks. Since BT system lacks a strong mechanism to identify peers’ authenticity, a large number of peers owned by one or several hosts may be forged, which can be exploited by attackers to perform advanced attacks. Work [17] gave a detailed study on Lying Piece Possession attack and Sybil attack. A malicious peer is able to employ the BitField and Have messages to announce arbitrary pieces that it actually does not own, which will induce the local rarest policy to make wrong piece selection and result in the unbalances in the amount of replicas for each piece. The Sybil attack tries to eclipse the connections of legitimate peers. The simulation results show that BT is susceptible to them.

III. Valid Connection Model Based on Leechers Attack

In this section, we briefly introduce the basic steps in Leechers attack. First, the attacker has to forge many identities because BT system has employed incentive mechanism to fight against selfish peers, which means a leecher cannot obtain data continuously from others if it denies to upload any blocks. Second, the attacker uses these identities to connect with legitimate users, and tells them that each identity is a new peer (All-0 Bitfield). When legitimate users randomly unchoke upload slots, these forged identities have opportunity to obtain valid data from them. Therefore, the attacker can obtain continuous data as long as there are enough identities. Finally, these identities controlled by the attacker collaborate with each other to request different parts of the resources until the resources are downloaded locally. The Leechers attack is illustrated in Fig. 1.

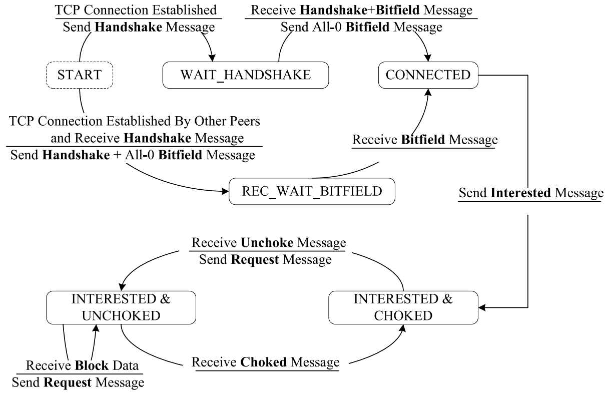
TCP Connection Established
Send **Handshake** Message
Receive **Handshake**+**Bitfield** Message
Send All-0 **Bitfield** Message
START
WAIT_HANDSHAKE
CONNECTED
TCP Connection Established By Other Peers
and Receive **Handshake** Message
Send **Handshake** + All-0 **Bitfield** Message
Receive **Bitfield** Message
REC_WAIT_BITFIELD
Send **Interested** Message
Receive **Unchoke** Message
Send **Request** Message
INTERESTED & UNCHOKED
INTERESTED & CHOKED
Receive **Block** Data
Send **Request** Message
Receive **Choked** Message

Fig. 1: The state diagram of Leechers attack

To analyze the damage of the Leechers attack, we use the following notations: n is the number of legitimate peers in a swarm, m is the number of upload connections of a legitimate peer, and k is the number of malicious peers (attackers) in a swarm.

A connection is considered as valid if the connection transmits actual data. The number of valid connections is $m \times n$. In normal situation, all connections are set up among legitimate peers and are valid. When a swarm is under attack, some valid connections are occupied by malicious peers and the bandwidths of these connections are wasted. In the text followed, we study the distribution of valid connections between legitimate and malicious peers.

Denote $F(n, m, k, t)$ as the number of valid connections obtained by legitimate peers under a Leechers attack after the t^{th} round optimistic unchoking. Denote $P(n, m, k, t)$ as the ratio between the number of valid connections after attack and that before attack.

Theorem 1. *The number of valid connections obtained by legitimate peers after the t^{th} round optimistic unchoking is*:

$$F(n,m,k,t) = \frac{k \times n \times (m-1)^t + n^2 \times m^t}{(n+k) \times m^{(t-1)}}, \quad t > 0 \quad (1)$$

Proof. When there is no attack ($t = 0$), i.e., there is no malicious peer ($k = 0$), the number of valid connections from the set of legitimate peers equals to that in the swarm, i.e.,

$$F(n,m,0,0) = m \times n, \qquad P(n,m,0,0) = 1 \qquad (2)$$

When the swarm is under attack, the total number of peers is $n + k$. Each legitimate peer releases a valid connection to perform optimistic unchoking, so that in total n valid connections are released. The probability of obtaining a valid connection by a legitimate peer is close to $\frac{n}{n+k}$. The number of valid connections obtained by legitimate peers after the first round is

$$\begin{aligned} F(n,m,k,1) &= m \times n - n + \frac{n}{n+k} \times n \\ &= F(n,m,0,0) - n + \frac{n}{n+k} \times n \quad (3) \end{aligned}$$

where $F(n,m,k,1)$ has three parts: $F(n,m,0,0)$ is the number of valid connections in the first round, n is the number of valid connections released from legitimate peers, and $\frac{n}{n+k} \times n$ is the number of valid connections relocated to legitimate peers. Next, we re-write Eqns. (3) in a recursion form as

$$\begin{aligned} F(n,m,k,1) &= F(n,m,0,0) - n + \frac{n}{n+k} \times n \\ &= F(n,m,0,0) \times \frac{m-1}{m} + \frac{n}{n+k} \times n \quad (4) \end{aligned}$$

The number of valid connections obtained by legitimate peers after the second round is

$$\begin{aligned} F\big(n,m,k,2\big) &= F\big(n,m,k,1\big) - \frac{F\big(n,m,k,1\big)}{m \times n} \times n \\ &\quad + \frac{n}{n+k} \times n \\ &= F\big(n,m,k,1\big) \times \frac{m-1}{m} + \frac{n}{n+k} \times n \quad (5) \end{aligned}$$

Generally, the number of valid connections obtained by legitimate peers after the t^{th} round is

$$\begin{aligned} F\big(n,m,k,t\big) &= F\big(n,m,k,t-1\big) \times \frac{m-1}{m} + \frac{n}{n+k} \times n \\ &= \frac{k \times n \times (m-1)^t + n^2 \times m^t}{(n+k) \times m^{(t-1)}}, t > 0 \quad (6) \end{aligned}$$

Theorem 2. *The ratio between the number of valid connections after attack and that before attack after the t^{th} round optimistic unchoking is:*

$$P(n,m,k,t) = \frac{k}{n+k} \times (\frac{m-1}{m})^t + \frac{n}{n+k}, \quad t > 0 \qquad (7)$$

Proof. The number of valid connections before attack follows $F(n,m,0,0) = m \times n$, and the ratio between the number of valid connections after attack and that before attack is

$$\begin{aligned} P\big(n,m,k,t\big) &= \frac{F\big(n,m,k,t\big)}{F\big(n,m,0,0\big)} = \frac{k \times (m-1)^t + n \times m^t}{\big(n+k\big) \times m^t} \\ &= \frac{k}{n+k} \times \left(\frac{m-1}{m}\right)^t + \frac{n}{n+k}, t > 0 \quad (8) \end{aligned}$$

From Eqns. (1) and (7), we can see that $F(n,m,k,t)$ and $P(n,m,k,t)$ depend on several factors, including the number of malicious peers k, the number of legitimate peers n, the number of upload connections of each legitimate peer m, and the number of round t. When m goes to infinitive, $P(n,m,k,t)$ approaches to one. This means that if there are a large number of upload connections, then the Leechers attack has little impact. When the number of round t goes to infinitive, $P(n,m,k,t)$ approaches to $n/(n+k)$, which is the ratio between the number of legitimate peers and that of all peers (legitimate plus malicious).

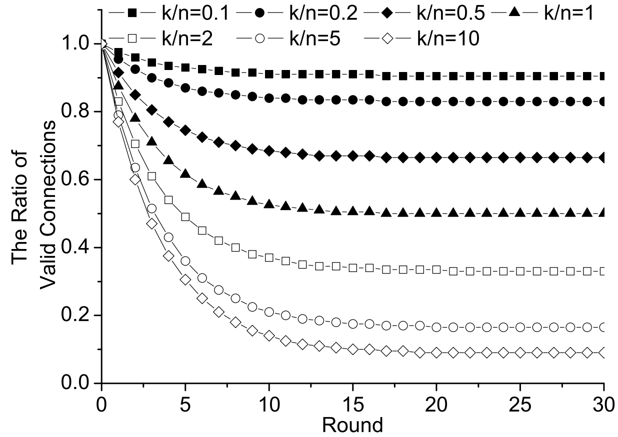

Fig. 2: The ratio of valid connections with different ratios of malicious peers vs. legitimate peers

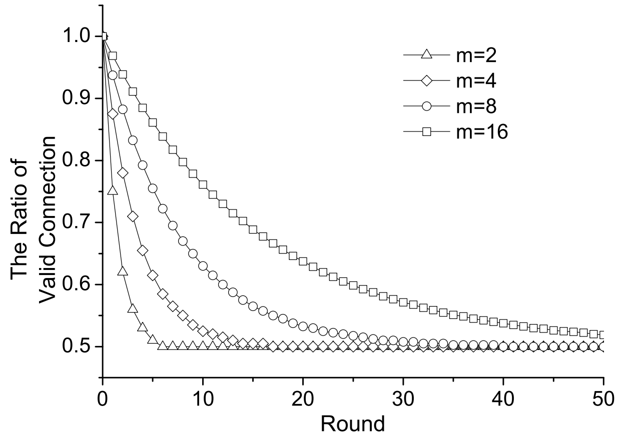

Fig. 3: The ratio of valid connections with different number of upload connections

Figs. 2 and 3 show the impacts of k/n and m on the ratio of valid connections. Fig. 2 shows the ratio for different values of k/n and a default upload connections value ($m = 4$). We can see that, the more malicious peers, the faster the ratio decreases. After 15 rounds, the curves become stable, which shows that the ratio of valid connections is determined by the ratio of malicious peers and legitimate peers. Fig. 3 plots the ratio for different values of m when $k/n = 1$, i.e., the number of malicious peers is the same as that of legitimate peers. Fig. 3 shows that, the more upload connections, the more valid connections are obtained by legitimate peers.

IV. Experimental Results

In addition to the theoretical evaluation given in Section III, we investigate the Leechers attack in real-world Internet.

A. *Experiments*

We set up both public and semi-public environments to evaluate the impacts of Leechers attack. The public environment consists of several servers with public IPs that can be connected directly from outside. Each server hosts thousands of lightweight malicious clients behaving like leechers. Besides, two legitimate testing clients are also deployed in the same network. To compare the difference between attack and non-attack scenarios, we install a filtering (blacklist) function into one of testing clients. The filtering function is able to reject attacking connections from malicious clients. However, it is improper and illegal to directly attack public torrents. In order to analyze the impacts in a more controlled environment, we also design a semi-public environment that includes a private tracker that can be accessed locally, and several servers with internal IPs that cannot be connected from outside. As a result, malicious clients are constrained and can only be connected by our testing clients.

Considering that the attack may cause a long time to download the resource, we define a ratio to estimate the duration even though the downloading has not been completed:

$$T_D = \frac{Duration}{DownloadedPercentage} \tag{9}$$

where *"Duration"* is the download time recorded by hour in the test, and *"Downloaded Percentage"* is the percentage of the resource that has been downloaded. The equation can be used to estimate the rest that has not been finished yet.

Table I presents the experimental results obtained from a real network and gives the download time in Leechers attack, in which malicious peers are almost two times as high as the number of legitimate peers. According to theoretical analysis, $P(n, m, k, t)$ will be $1/3$ in Eqn. (7) when t goes to infinitive, indicating that the time under attack is extended to at most three times. The experiment over real-world Internet shows about 2.75 times ($1.65/0.6$), which matches well with our proposed theoretical model.

However, we also observed a wide fluctuation of download time, especially using different kinds of BT clients to download. Actually, there are other factors to influence Leechers attack potentially, which has not been modelled in our work yet. The possible reasons can be inferred as follows:

- As an open protocol, BT does not give the specifications on how to implement a client. To have a higher market share, some developers have made their own efforts in improving client performance. Due to the selfish and low-rate peers generated by Leechers attack, progressive clients can evaluate the transmission rate of different peers and replace slow peers with others to get a better performance. For example, if uTorrent [18] finds there is no activity in a connection, the client will stop it. On the other hand, BitSpirit [19] does nothing to the slow connections.
- Due to the dynamic nature of peers, the ratio between the number of malicious peers and the number of legitimate peers continues changing over time.

TABLE I: Comparison of Download Time

Name	TV1	
Size	191MB	
Scenarios	Non-attack	Attack
Ratio(malicious/legitimate)	0	≈ 2
T_D	0.6	1.65

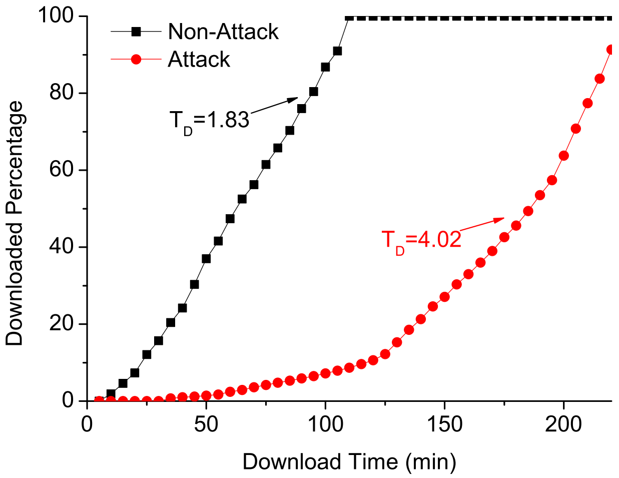

Fig. 4: The Downloading in BitSpirit

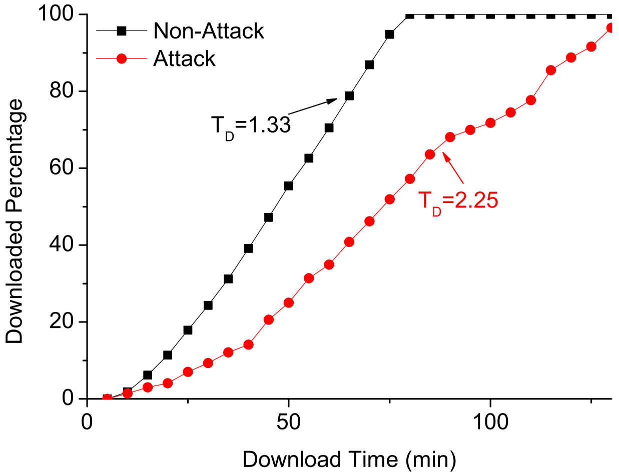

Fig. 5: The Downloading in Vuze

- Different connections have their own transmission rates. The high-bandwidth connections built amongst legitimate users can send more data per unit time, which will decrease the effects of Leechers attack. However, our model does not consider the difference of connections.

Therefore, another set of experiments were performed to demonstrate the impacts of Leechers attack on different BT clients. Three popular mainline clients are chosen, including BitSpirit [19], Vuze [20] and uTorrent [18], whose results are shown in Figs 4, 5 and 6 respectively. In the experiments, malicious peers are about twice as many as legitimate ones. From these figures, it is shown that compared with normal situation the downloading of BitSpirit is delayed the most

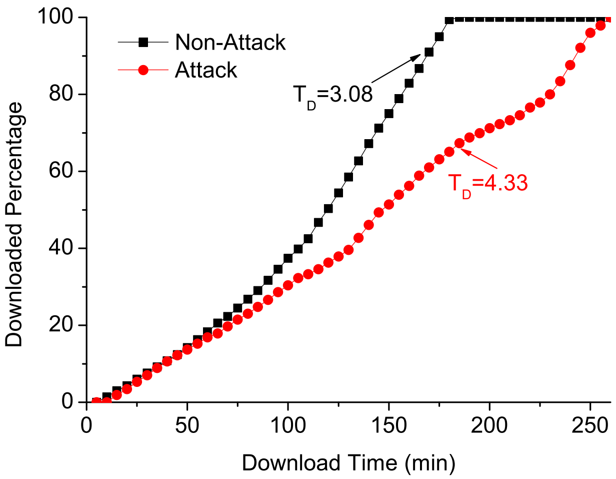

Fig. 6: The Downloading in uTorrent

while uTorrent is delayed the least. The results indicate that BitSpirit is more susceptible to Leechers attack and uTorrent is more resistant.

B. Discussion

Through the theoretical modeling and real-world experiments, it is shown that Leechers attack on BT system has the characteristic of resource confrontation. System efficiency is closely related to the percentage of legitimate users and the number of upload slots in our model. Therefore, there are two effective ways advised in this paper:

- **Limiting the number of malicious users**. The attacker exploits the vulnerability of authentication to generate thousands of identities without any constraint at a small cost. These identities are used to take up many valid connections to benefit from legitimate users, but they do not contribute any data back to the swarm. The more malicious users are, the less valid connections can be obtained by legitimate users. Thus, to limit the number of malicious users, it is necessary to strengthen the process of authentication or detect malicious users. The key of identity authentication is to guarantee the strict relationship between the identity and physical machine, for example, using unique physical identifier to differentiate users. The detection of malicious users can be achieved by evaluating the transmission rate or the global contribution to the swarm to tell malicious users from legitimate ones. However, this method has to modify the existing protocol and implementation of BT system or clients, which may be not impractical.
- **Increasing upload slots of legitimate users**. According to our analysis, the effects of Leechers attack can be decreased by increasing the number of legitimate users and their upload slots. However, the number of legitimate users mainly depends on the popularity of resources that is hard to control. Thus, it is a simple and effective way to increase upload slots of legitimate users, which also minimizes the modification of current implementation.

V. Conclusion

In this paper we presented a valid connection model to theoretically analyze the Leechers attack and estimate the distribution of valid connections amongst malicious and legitimate users. The analytical model indicates that the effect of the Leechers attack is closely related to the number of malicious peers, the number of legitimate peers and the number of upload slots of each legitimate peer. Our model can predict the duration of downloading process under the Leechers attack. We conducted real-world experiments in the Internet. The experimental results match well with our model, which demonstrates the correctness of our model.

Acknowledgment

This research is supported by the National Basic Research Program (973 Program) of China under Grant No. 2011CB302605, the National High-Tech Development 863 Program of China under grants No. 2011AA010705, 2012AA012506 and 2010AA012504, the National Natural Science Foundation of China under grants No. 60903166, 60203021 and 61173145, and by the US National Science Foundation (NSF) under grants CNS-0963578, CNS-1022552, and CNS-1065444.

References

[1] BitTorrent. [Online]. Available : http://www.bittorrent.org/
[2] B. Wei, G. Fedak, and F. Cappello, "Collaborative Data Distribution with BitTorrent for Computational Desktop Grids," *in ISPDC'05*, pp. 250-257, 2005.
[3] N. Parvez, C. Williamson, A. Mahanti, and N. Carlsson, "Analysis of bittorrent-like protocols for on-demand stored media streaming," *in SIGMETRICS'08*, pp. 301-312, 2008.
[4] J. Falkner, M. Piatek, J. P. John, A. Krishnamurthy, and T. Anderson, "Profiling a million user dht," *in IMC'07*, pp. 129-134, 2007.
[5] P. Dhungel, D. Wu, B. Schonhorst, and K. W. Ross, "A measurement study of attacks on BitTorrent leechers," *in IPTPS'08*, pp. 7-7, 2008.
[6] K. E. Defrawy, M. Gjoka, and A. Markopoulou, "BotTorrent: misusing BitTorrent to launch DDoS attacks," *Proc. of the 3rd USENIX workshop on Steps to reducing unwanted traffic on the internet*, 2007.
[7] N. Liogkas, R. Nelson, E. Kohler, and L. Zhang, "Exploiting BitTorrent for fun (but not profit)," *In IPTPS'06*, 2006.
[8] M. Sirivianos, J. H. Park, R. Chen, and X. Yang, "Free-riding in BitTorrent networks with the large view exploit," *In IPTPS'07*, 2007.
[9] Mininova. [Online]. Available: http://www.mininova.org/
[10] BT@China Union. [Online]. Available: http://bt.btchina.net/
[11] E. Rescorla, "Introduction to Distributed Hash Tables," *IETF-65 Technical Plenary*, Mar. 2006.
[12] Peer Exchange. [Online]. Available: http://en.wikipedia.org/wiki /Peer_exchange
[13] Cohen, "Incentives build robustness in BitTorrent," *Proc. of the 1st Workshop on Economics of Peer-to-Peer Systems*, 2003.
[14] S. Jun and M. Ahamad, "Incentives in BitTorrent induce free riding," *Proc. of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*, pp.116-121, 2005.
[15] T. Locher, P. Moor, S. Schmid, and R. Wattenhofer, "Free Riding in BitTorrent is Cheap," *Proc. of HotNets-V*, pp. 85-90, 2006.
[16] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "Do incentives build robustness in BitTorrent?" *Proc. of the 4th USENIX Symposium on Networked Systems Design & Implementation*, pp.1-14, 2007.
[17] M. A. Konrath, M. P. Barcellos, and R. B. Mansilha, "Attacking a Swarm with a Band of Liars: evaluating the impact of attacks on BitTorrent," *in P2P'07*, pp. 37-44, 2007.
[18] uTorrent. [Online]. Available : http://www.utorrent.com/
[19] BitSpirit. [Online]. Available : http://www.bitspirit.cc/
[20] Vuze. [Online]. Available : http://www.vuze.com/