ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 2, pp: (153-157), Month: April-June 2014, Available at: www.researchpublish.com

Content Poisoning in Peer to Peer Network

[1]Pratibha Singh, [2]Rishi Srivastava

Computer Science
Babu Banarasi Das University

Abstract: **Now a day's poisoning attack are very common in peer-to-peer (P2P) networks. In this condition of poisoning refer corrupt or infected content which share by malicious peer and system destabilize attempt and network waste the bandwidth. In content sharing system in P2P network are highly vulnerable to content poisoning. We are trying to intrusion this distribution of the files, recently much attention has attracted by the content poisoning. Although the aims of the content poisoning blackout users by splitting in P2P networks by the poisoning chunks. Several anti P2P companies have tried method such as pollution or index poisoning .The method of pollution is decrease target files availability, in P2P sharing file network by splitting of duplicate or dummy files. In this paper, we propose a strategy to minimize the threat of content poisoning, while requiring less verification overhead on the peers participating in the network.**

Keywords: **Peer to Peer, content poisoning, falsification, probabilities, verification.**

I. INTRODUCTION

Peer to peer (P2P) networks have occurred as a one of the very popular in networks to share files in P2P networks system. P2P directly related to the sharing file of music, videos etc. and occurs frequently illegally to avoid the illegality of the content in P2P networks, some P2P companies introduced some types of attack. P2P networks have a distribution nature but there have a lack of security issues of authority of the central networks system, the main reason of the success P2P network system. P2P is more valuable to the malicious attack into the network. Such attacks are free- riding, collusion, denial of service, distributed denial of service, etc.

In content poisoning introducing some massive type of data to be injected and genuine data become corrupt. They might be a white noise, low quality, and corrupt, virus and infected type of data is injected in this. [1]

In some studies of P2P networks system are indicates that afflicted to the content poisoning. Some popular songs which are share in the network of the P2P found content poisoning mostly, most of the songs are polluted from the poisoners. In order file illegal distribution relegate some anti P2P companies have to introduced method that method is known as index poisoning or content pollution[2]. Content poisoning decreasing the quality of the data availabity and create more pollution in the P2P networks to avoid the content poisoning we propose method who identified the pollution in the networks and probabilistic method to detect the content poisoning.

II. RELATED WORK

The problem of the content and index poisoning, many several researchers have been proposed the solution of the problems, overlay networks especially in DHT based. Deliver contents are not require many expensive servers in P2P networks. [1]

In P2P file sharing network, pollution are divided into following types, that is pollution and the second is index poisoning. In P2P file sharing networks, pollution is more valuable to interrupt the diffusion of the desired files. Content pollution is very popular form in P2P file sharing Networks. Digital recording (mp3 (audio), mp4 (videos)) of target party in content pollution. It's a method to reduce the availabity of desired files by the spreading of dummy files in P2P sharing networks. A dummy has all most same to genuine file but its content is forged. Such as noisier, corrupt files, or inserting another file in middle. We observed that insert undecodable white noise into the middle of the song. In content pollution attack, make target content inoperative by the attacker to the help of changing content in another regardless content. In large amount of content pollution available for sharing in network. We are unable to differentiate between polluted and unpolluted files, the unsuspicious client transfer the polluted file into their self maintain file-sharing content; other client may also download from the polluted files [3].

In attack of index poisoning inject the massive the bogus data into the target location set for index. Any user searched a data from the target file then it the result of the index returns as a bogus data or massive data, fake location (IP address, port no., service port no.). As we know that index poisoning this highly vulnerable in file sharing of both the system (structured and unstructured system).

While incentives are very useful at depressing self-recentness, curtailing misconduct requires the ability to punish spiteful peers. As deliberated on staring, the reputation system of primary functions is to inform agents as to which peers are likely to defect on a transaction. Not only does adversary avoidance benefit well-behaved peers, but who will quickly unable disseminate bad resources or cheat to the other peers malicious punished. E-commerce sites, such as eBay use reputation systems not only to provide good customers information on sellers giving buyers a sense of security, but also to discourage misbehaviour in the first place [7].

III. PROPOSED METHOD

A. Detecting Pollution in P2P Sharing Networks:

1. The pollution index falsification:

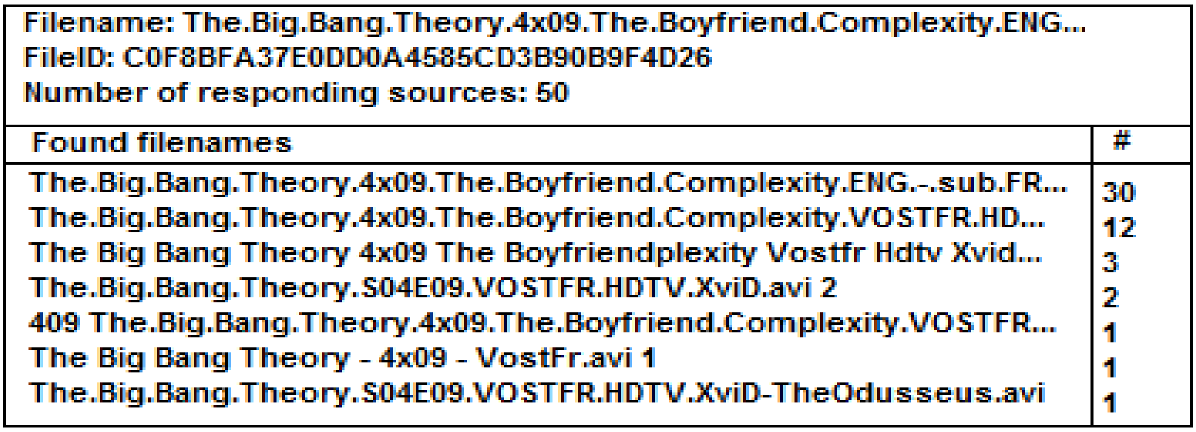
We emphasize and compute pollution new form which is spread into the network. While many unexcited files in index poisoning advertises which we can't be downloaded. Advertising a single file with many different file names consists in index falsification, and subsequently many different keywords. Which are not related to the real content? Each false filename is preciously made popular to the help of polluters.

This form of the pollution is more dangerous because it leads to user's undesirable content download, it's nothing but it's just a waste of the network resources process and for the safety of the user the downloaded file could be harmful. The downloaded content can be a video hurting users feeling (low quality, pornographic content) or a malware. Many false positives create by this pollution when monitoring these illegal contents to any network or the fact any users never monitoring this illegal files content. On the daily basis when we suffer from the pollution of this. It's very important to study to investigate these problems.

2. Approach for the detection of the index falsification*:*

In index falsification to detection this pollution, which might to be collect all the different file-names attached to a file and find out their consistency. Still, it's making very tough to retrieve the scheme of the double indexation scheme search from a keyword, it is possible that we can obtain the different keyword linked and their details (file size, file name, etc). But from a keyword search collected all the different files include the genuine keyword in their file-name, if sometime some files which is not related indexed through keywords in the DHT. On the other hand when we search from the source, it's a possibility to obtain the file from all the sources. So it's not matter that which sources uses the file-name. However, at the level of DHT is not an important published? The second level of DHT the possibility of filenames can't be obtained by regular level lookups [1].

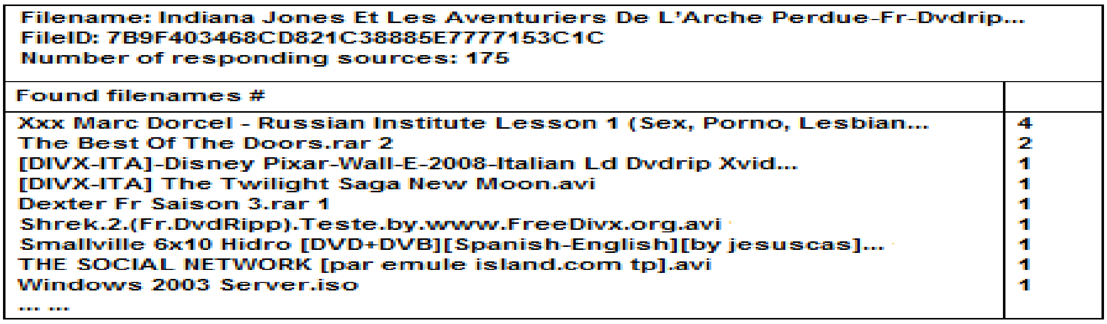
Table 1 - Example of consistent filenames retrieved from the responding sources for a clean file



Filename: The.Big.Bang.Theory.4x09.The.Boyfriend.Complexity.ENG...
FileID: C0F8BFA37E0DD0A4585CD3B90B9F4D26
Number of responding sources: 50

Found filenames	#
The.Big.Bang.Theory.4x09.The.Boyfriend.Complexity.ENG.-.sub.FR...	30
The.Big.Bang.Theory.4x09.The.Boyfriend.Complexity.VOSTFR.HD...	12
The Big Bang Theory 4x09 The Boyfriendplexity Vostfr Hdtv Xvid...	3
The.Big.Bang.Theory.S04E09.VOSTFR.HDTV.XviD.avi 2	2
409 The.Big.Bang.Theory.4x09.The.Boyfriend.Complexity.VOSTFR...	1
The Big Bang Theory - 4x09 - VostFr.avi 1	1
The.Big.Bang.Theory.S04E09.VOSTFR.HDTV.XviD-TheOdusseus.avi	1

Table 2 Example of consistent filenames retrieved from the responding sources for a clean file



Filename: Indiana Jones Et Les Aventuriers De L'Arche Perdue-Fr-Dvdrip...
FileID: 7B9F403468CD821C38885E7777153C1C
Number of responding sources: 175

Found filenames #	
Xxx Marc Dorcel - Russian Institute Lesson 1 (Sex, Porno, Lesbian...	4
The Best Of The Doors.rar 2	2
[DIVX-ITA]-Disney Pixar-Wall-E-2008-Italian Ld Dvdrip Xvid...	1
[DIVX-ITA] The Twilight Saga New Moon.avi	1
Dexter Fr Saison 3.rar 1	1
Shrek.2.(Fr.DvdRipp).Teste.by.www.FreeDivx.org.avi	1
Smallville 6x10 Hidro [DVD+DVB][Spanish-English][by jesuscas]...	1
THE SOCIAL NETWORK [par emule island.com tp].avi	1
Windows 2003 Server.iso	1
... ...	

3. Comparison metric for pollution detection:

A file-ID given, determine if the file is reliable or polluted by the index falsification. Our detection the different file names given by the sources is based on overall consistency. To calculate two filenames similarity, we are using matrix to evaluate the similarity their set of keywords. Let us P and Q be a keywords sets. Where keywords associated with the desired file-name is being P and keywords associated with a file name regained from the sources are being Q. Here we using Tversky index [13] is a metric similarity (where $\alpha = \beta$) used in mining of data and defined by:

$$K(P,Q) = \frac{|P \cap Q|}{|P \cap Q| + \alpha * |P - Q| + \beta * |Q - P|} \quad (1)$$

(P, Q) ε [0, 1] and more accurately 1 returns if the both files have the same name and if they return 0 means that there have no common keywords.

Now we define for each file pollution coefficient S for P as an average function of coefficient similarity for all file-name Q*i* which is regained from the sources n. Fig 1

$$S(P) = 1 - \frac{\sum_{i=1}^{n} K(P,Q)}{n} \quad (2)$$

$$K\ (P,\ Q) = \frac{|P \cap Q|}{|P \cup Q|} \quad (3)$$

Now we defined the probability check $S\ (P_c)$. In peers of the network adopt a method which verification to reduce the Pc value and it's maintain from the behaviours of the past content downloaded of the such peers k. in future from peers k represent the probability and maintain a list of verification Fig 2.

IV. RESULT

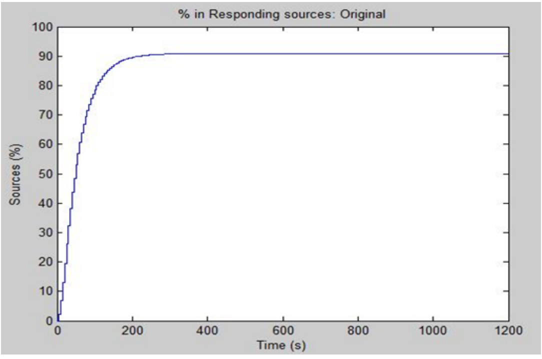


Fig.1 % of the responding sources discovered

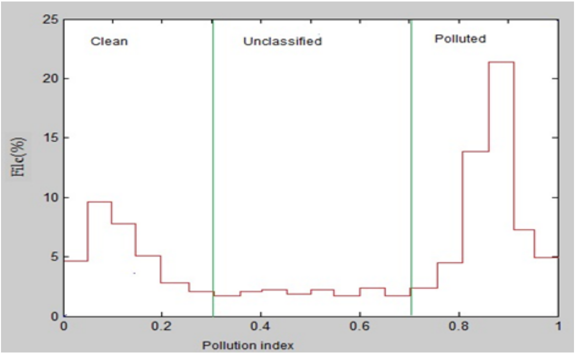


Fig.2 Distribution of files according

We introduce an algorithm who reduces the poisoning control. As soon as probability checks value is increase that means this peers mostly time shared infected file to the peer of the network. Support from a peer x shared files and other peer download the files but mostly time peer x share infected files then they increase its probability check cost and other when its mostly time share genuine files to the peers then the value of check probability has to decrease $S(P_c^k)$.

We introducing four type of verification, first verification are no verification, and second verification is probabilistic verification, third is dynamic verification and last is full verification.

Proposed Algorithm:

Initialize $S(P_c^k)$ to 1.0 for all peer k.

```
For the file F from peer k.
With probability check S$(P_c^k)$ verify file F.
{
If
File is infected.
$S(P_c^k)$ ← (0.7~1.0).
Then delete file.
Else
Keep and share.
}
Otherwise
Keep and share.
```

In no verification P_c=0.0 there have no verification overhead is present that types of peers never verify files there have possible maximum infected files are presents. In probabilistic verification the probability check is P_c=0.5 there have fifty-fifty change to verify the file. In this file could be infected or be verified. In dynamic verification is the collection of full and probabilistic verification means that in this file verification its possibility to full verifies or they have a possibility of fifty-fifty. In full verification P_c =1.0 means that their peers will always verify the peers. There have full verification overhead and no infection.

The performance shows in fraction verifications for tasted four situations vs. simulation time, and performs fraction infections which are occurred in P2P network system vs. time. This fraction depends on total downloaded calculation fraction based files.

The situations for full verification and no verification basically what we presume and our simulations validate: when we verified 100% file then 0% infection occurs in our P2P network system. Basically we are trying to exclude cost of verification then cost reaches maximum rate as possible. The exact value of the injected poisoned content (20%). This is almost same to the related of prisoners in P2P network system.

In our proposed algorithm is tried to reduce the verification above as much as possible that we reduce.

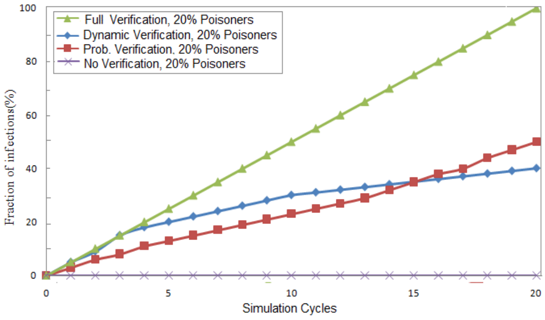


Fig 3. Fraction of the verification in the care of 20% poisoners sharing infected files

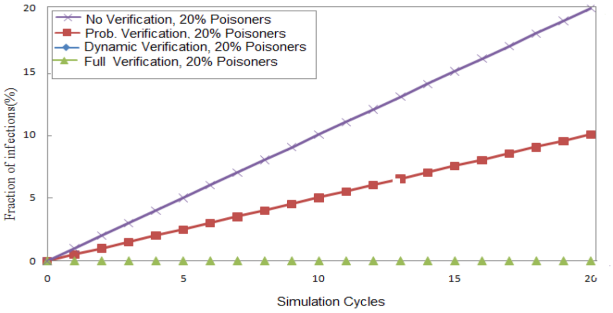


Fig 4. Fraction of the infections in the care of 20% poisoners sharing infected files

V. CONCLUSION

The content poisoning problem is the most important threat for any P2P system. It prevents the system from working efficiently. As a result, some researchers proposed several mechanisms to provide verification probability check to the effect of content poisoning in a P2P system or network. Here we have proposed the verification mechanism. Which can prove quite effective in this regard? This work also proposes the method of detection which can also prove quite useful in reducing the activity of content poisoning. In our work, behavior factor is the most important parameter for analyzing the behavior of content poisoning.

The approach that we have used here controls the content poisoning activity and encourages the peers to be best in the network. It forces the peers to become contributor rather than consumers. Our work is an improvement over many other previous approaches. The results show that as the number of nodes increases in the network, number of content poisoning decreases in the network. It also depicts that behavior factor is indirectly proportional to verification probability check, i.e. if the behavior factor of a node increases, the probability check will decrease and vice-versa.

VI. FUTURE WORK

In our future works, we will investigate the polluting behaviors in order to understand precisely how this pollution is achieved. Then, we will design a detection mechanism which can operate earlier in the download process to avoid the initialization of many connections towards the responding sources. Our solution will also need to be suitable for real implementations (by keeping backward compatibility and minimizing the overhead) in order to protect current P2P networks.

REFERENCES

[1] Content Pollution Quantification in Large P2P networks : a Measurement Study on KAD: Guillaume Montassier, Thibault Cholez, Guillaume Doyen, Rida Khatoun, Isabelle Chrisment*, Olivier Festor** (IEEE P2P'11) (2011)

[2] On Combating Content Poisoning in Peer-to-Peer Networks: Mohammed Hawa, *Member*, *IAENG*, Raed Al-Zubi, Khalid A. Darabkh, and Ghazi Al-Sukkar** July 3 - 5, 2013, London, U.K.

[3] Interoperability of Peer-To-Peer File Sharing Protocols: Siu Man Lui and Sai Ho Kwok ** Categories and Subject Descriptors: D.2.11. [Software] : Software Engineering – Software Architecture, D.2.12

[4] Controlling File Distribution in The Share Network Through Content Poisoning : Masahiro Yoshida †, Satoshi Ohzahata ∗, Akihiro Nakao §, Konosuke Kawashima** 2010 24th IEEE International Conference on Advanced Information Networking and Applications

[5] Napster Website:http : / /www.napster.com.

[6] Wikipedia.

[7] Prevention of Index-Poisoning DDoS Attacks in Peer-to-Peer File-Sharing Networks*: Xiaosong Lou, Student Member IEEE and Kai Hwang, Fellow IEEE**In 2005.

[8] Cristiano Costa and Jussara Almeida. Reputation systems for fighting pollution in peer-to-peer file sharing systems. In P2P '07: Proceedings of the Seventh IEEE International Conference on Peer-to-Peer Computing.

[9] Thibault Cholez, Isabelle Chrisment, and Olivier Festor. Efficient DHT attack mitigation through peers' ID distribution. In Seventh International Workshop on Hot Topics in Peer-to-Peer Systems - HotP2P 2010, Atlanta USA, 04 2010. IEEE International Parallel & Distributed Processing Symposium.

[10] C. Costa and J. Almeida. "Reputation systems for fighting pollution in peer-to-peer file sharing systems," *In:* Proceedings of the Seventh IEEE International Conference on Peer-to-Peer Computing, 2007, pp. 53–60.

[11] C.-L. Hu and Z.-X. Lu, "Downloading trace study for BitTorrent P2P performance measurement and analysis," *Peer-to-Peer Networking and Applications*, Volume 5, Issue 4, 2012, pp. 384–397.

[12] J. Shneidman and D.C. Parkes, "Rationality and Self interest in Peer-to-Peer Networks",

[13] Amos Tversky. Features of similarity. In Psychological Review, volume 84, pages 327–352, 1977.