Content Poisoning in Peer to Peer Networks

A Thesis Submitted
In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF TECHNOLOGY

In

Computer Science & Engineering (Computer Networks)

By

PRATIBHA SINGH
Enrollment No.: 11204471044

Under the Supervision of

Mr. RISHI SRIVASTAVA
Department of Computer Science
BBDU, Lucknow

To the
School of Engineering

BABU BANARASI DAS UNIVERSITY
LUCKNOW
May 2014

DECLARATION

I hereby declare that this submission is my own work and that, to the best of my knowledge an belief, it contains no material previously published or written by another person or material which to a substantial extent has been accepted for the award of my other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature

Name: Pratibha Singh
Enrollment No.: 11204471044
Roll No.: 1120447014

CERTIFICATE

It is certified that the work contained in this thesis entitled **"Content Poisoning in Peer to Peer Networks"** by PRATIBHA SINGH (Roll No 1120447014), for the award of Master of Technology in Computer Science and Engineering from Babu Banarasi Das University has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

Signature:

Date:

Mr. Rishi Srivastava
Sr. Lecturer
Department of Computer Science
BBDU, Lucknow

Content Poisoning in Peer to Peer Networks

Pratibha Singh

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.

Index terms—Peer to Peer, content poisoning, falsification, probabilities, verification.

ACKNOWLEDGEMENT

I would like to take this opportunity to express my deep sense of gratitude to all who helped me directly or indirectly during this thesis work.

Firstly, I would like to thank my supervisor, Mr. Rishi Srivastava, for being a great mentor and the best adviser I could ever have. His advice, encouragement and critics are source of innovative ideas, inspiration and causes behind the successful completion of this dissertation. The confidence shown on me by him was the biggest source of inspiration for me. It has been a privilege working with him from last one year.

I am highly obliged to all the faculty members of Computer Science and Engineering Department for their support and encouragement. I also thank our Dean Prof. (Dr.) Seethalakshmi K. and H.O.D. (CSE) Dr. Reema Srivastava for providing excellent computing and other facilities without which this work could not achieve its quality goal.

I would like to express my sincere appreciation and gratitude towards my BBD University friends for their encouragement, consistent support and invaluable suggestions at the time I needed the most.

Finally, I am grateful to my parents for their support. It was impossible for me to complete this thesis work without their love, blessing and encouragement.

PRATIBHA SINGH

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Peer to Peer (P2P) is an alternative network model which provided by the traditional client server model. P2p networks use decentralized model in which each machine mention as a peer. In decentralized model a peer play a role of client and server at the same time. Peer can initiate the requests to other peers, incoming request at the same time from other peers of the networks. But the other hand in client server model, client only send the request to server and user wait for the server response.

Peer can upload and download at the same time. Peers are improves the increasing added number of the network. Peer can handle themselves into the ad-hoc network as to communicate. They work together and share bandwidth with each other for file sharing.

All p2p systems have common characteristics in them: resource sharing, decentralized and self organization. P2P networks have another characteristics that is, its capability in concept of fault tolerance. From the network of the P2P is disconnected or goes down, then other peer application will continue by using some other peers. For example, suppose a file sharing system (BitTorrent) any user or client downloading a files data are also serving as servers, when user supporting peer is not responding to the user then user searches for the other peer and picks up parts of the file where the old peer was, continue the downloading. Peer- to- peer networking, often referred to as P2P, is perhaps one of the most useful and yet misunderstood technologies to emerge in recent years. When people think of P2P they usually think of one thing: sharing music files, often illegally. This is because files sharing applications such as Bit Torrent have risen in popularity at a staggering rate, and these applications use P2P technology to work. Although P2P is used in file-sharing applications, that doesn't mean it doesn't have other applications. But we can see in the world around us that P2P can be used for a vast array of applications, and is becoming more and more important in the interconnected world in which we live.[5][6]

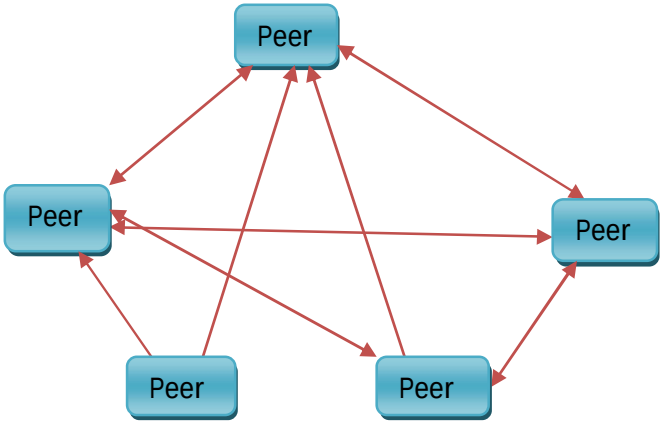


Figure 1.1:- Peer to Peer Networks

1.1.1 Overlay network:-

Overlay networks such as the Content Addressable Network (CAN), 2 Chord, 3 Pastry, 4 and Viceroy5 create a virtual topology on top of the physical topology. In this sense, TTL-based P2P networks are also a type of overlay, but we use the term here to refer only to networks that create virtual topologies based on node-content attributes. Some networks, such as Chord, organize the network on the basis of each participating node's IP address; other networks use the node's stored data as the organizing content [12].

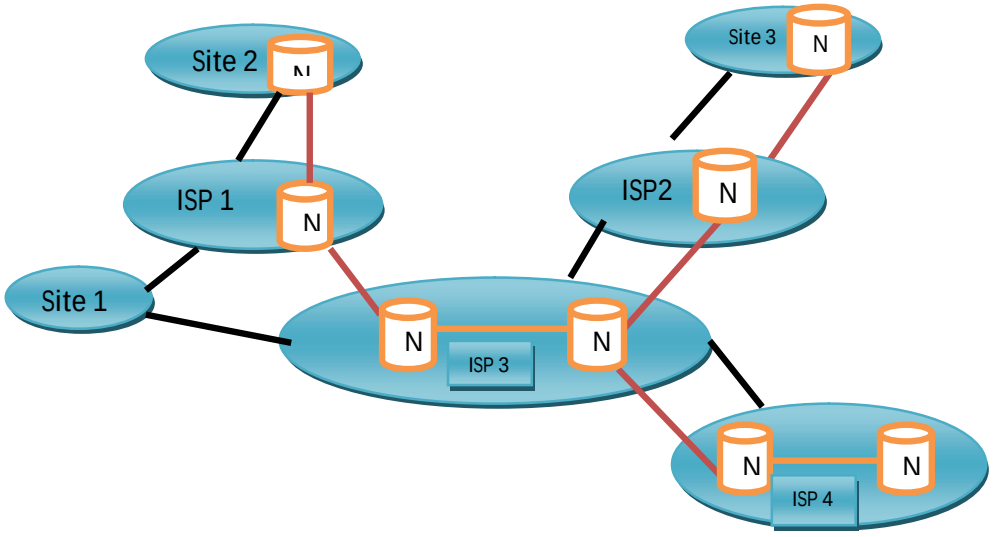
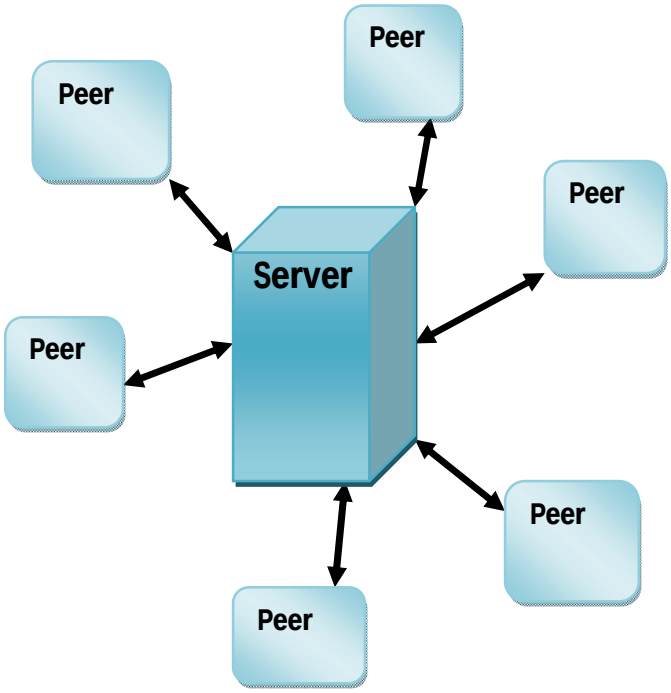


Figure 1.2:- Overlay network in P2P

1.1.2 Architectures for Peer-to-Peer network: -

Broadly, there are two different architectures for peer-to-peer networks: Centralized and Decentralized.

1.1.3 Centralized Peer-to-Peer systems: The centralized P2P systems consist of a central server, set of users and data base of files. Each user has a sub set of files and an access to all the files from all users in the systems. The request among all the users is communicated through the central server in order to retrieve the files. The central server, in turn, searches for the requested files in its database of files shared by other users. A list of matching files is created and sent back to the users. The user has the choice of selecting the desired files from the list and opening a direct connection to the other users. This connection helps in transferring the files from one peer to the other without storing it in the central server. Show the central server consists of only peer information and directory information of the shared files.



. Figure 1.3:- Centralized P2P Networks

Advantages:

- Consumes less network resources.
- Files can be found in lower cost.
- Files can be located relatively quickly and efficiently.
- Files transfer puts no load the server.

Disadvantages:

- Prone to central point of failure.
- Expensive to scale the central server.
- The central server might get over loaded.
- As the central server index is updated periodically, there is a possibility of receiving out dated information.

1.1.4 Decentralized P2P System: -

These systems consist of a distributed directory structure. These systems can also be classified into 2 types:

- **Structured P2P systems**: In these systems, overlay connections are fixed. It uses distributed hash table based indexing. Examples of systems which support distributed hash table functions are:-Tapestry, Pastry, Chord and Content Address Network. In these systems, contents are stored and retrieved according to strict rules. The current Node ID and content name of each node is hashed to a key. Content is stored in the node whose key is closest to the content key and query routing processes forward queries to neighboring nodes whose keys are closer to the query object keys.

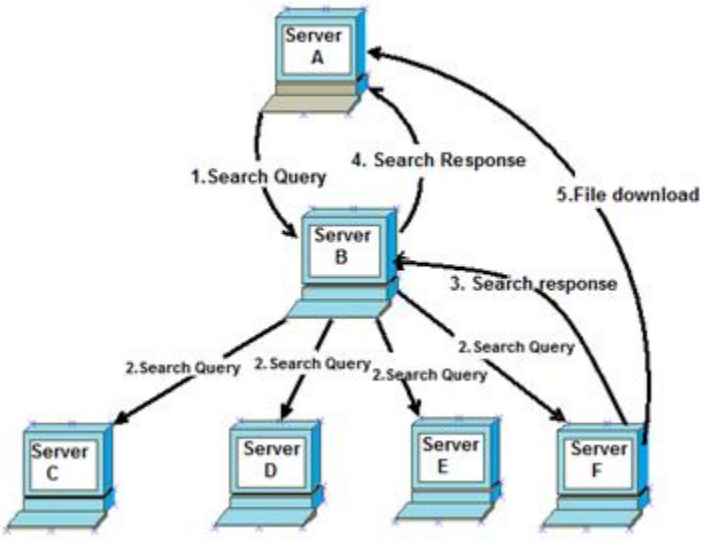


Figure 1.4: Decentralized Peer to Peer

ADVANTAGE:-

- Provides guarantee of finding data.

DISADVANTAGE:-

- No flexibility
- Keyword searching not possible.

- **UNSTRUCTURED P2P SYSTEMS**: -

There is no algorithm for organization of network connections in these systems. There are 2 types of unstructured systems: Pure and Hybrid Unstructured Systems. In pure decentralized P2P systems, peers are totally equal, e g. Gnutella. Here, peers are identified by their IP address. So, if any other peer wants to join the network, it must be familiar with at least one existing peer of the network. Incoming peer attempts to form a TCP connection with the known peer until the connection is set up with any existing peer of the network. In pure P2P systems, the query is done in flooding style where one asks their neighbors, who in turn ask their neighbors. Query is forwarded till TTL value exists. Once TTL value is exhausted, the query is dropped. If the query is hit, then the file is fetched directly between querying node and query hit node. In hybrid decentralized P2P systems, there is a super peer which acts as the main server for a set of clients. The clients are called Leaf Peers. They are attached to any super node which in turn is connected

to other super nodes. Leaf node queries the super node and it searches for the requested file in the peers attached to it. If the requested file is found, then the node having the file is given as input to the queried node and a TCP connection is established and the file is downloaded. Else the query is forwarded to the super node.

ADVANTAGE:

- Totally decentralized
- Less search cost
- Powerful search semantics

DISADVANTAGE:

- Search may be large
- Time taking
- High number of nodes

1.2 Motivation: -

Current approaches used for controlling content poisoning in peer-to-peer networks are either probabilistic or incentive based model. Some authors also have tried to eliminate the content poisoning problem of the peer-to-peer network by combining both the models. Some have proposed techniques on the basis of the traffic on the peer-to-peer network. It has been seen that nearly 70% of Gnutella users share no files, and nearly 50% of all responses are returned by the top 1% of sharing hosts. The major problem with content poisoning present in the system is that it degrades the performance of the whole system as all the queries will be directed towards few contributors present. This will make the contributors heavily loaded. When these contributors reach their threshold of handling a number of downloads then most queries from other peers will be roaming in the network. The other problem with content poisoning is that it creates vulnerabilities in the system like if there are only a few number of contributors present in the system and they are responsible for all the uploads in the system, then these few nodes will act as a centralized server failing our purpose of a decentralized network[5].

1.3 Problem to be analyzed: -

The common problem of the peer-to-peer network “the content poisoning” affects the performance of any of the peer-to-peer network. The previous approaches proposed do not take

accounts of treating the nodes according to their levels of the content poisoning. The mechanism used earlier only attracts the content poisoning to contribute to the system. Neither it can force them to share nor can it ban them from joining the network. Here in our dissertations we focus to make users bounded for sharing and increasing their online times.

To overcome the problem of content poisoning in the peer-to-peer network for improving the performance of the peer-to-peer network we have compared the scalar values of each peers and treated them accordingly.

In this dissertation, we have made an attempt to solve the mentioned problems:

1. Content poisoning control using scalar values of peers and the entire network.
2. limiting the search and download activity of the peers depending upon their behavior in the network.
3. Prohibiting the download from the upper priority level nodes.
4. Encourage the users for sharing much and rich contents.
5. Increase the online times of the users.
6. Decrease the query overhead on the contributor's node.

1.4 Organization of Report: -

This dissertation report comprises of five chapters including this chapter that introduces the topic and states the problem. The rest of the report is organized as follows. Chapter2 gives the background of content poisoning, description of the impact of content poisoning and brief literature review of related work. Chapter 3 describes the approach proposed for content poisoning control by the behavior of the peer in the network. Chapter 4 gives the details of experiments performed and discusses the performance reduction obtained and Chapter 5 concludes the dissertation work and gives suggestions for future work.

Chapter 2

Related Work and Background

2.1 Related work: -In real-life two types of poisoning exist

- Content Poisoning
- Index Poisoning

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]

P2P Networks	**Bit Torrent family[2]**	**Gnutella family[11]**	**eMule family[18]**
Chunking Scheme	Divide files into fix sized chunks (256KB), called pieces.	Peers negotiate the chunk size at run time, 6 KB chunks by default.	Divide into 9500-KB parts, each has 53(180 KB) chunks.
Hash Distributed	SHA hashing at pieces level, embedded in index file and distributed.	SHA hashing applied to entire file to generate a unique file ID, no chunk-level hashing.	MD-4 hashing at part level, peers exchange part level hash set to detect corrupted content.
Poison Resistance	Poison detected at pieces level, each is handle independently.	Poison detected only after download the entire file, heavy overhead if poisoned.	Poison detected at part level, works if part hastset is not poisoned.
Download Policy	Keep clean and discard poisoned pieces, repeated download until all chunks are clean	Repeat download entire until all chunks become clean, the most time –consuming policy	Keep clean discard poisoned parts, repeated download until all parte are clean
Example Networks	BitTorrent, snark, bitcomet, BNBT, bittyrant,etc	Gnutella,KaZaA, LimeWire, etc	eMule, \aMule, iMule, FastTrack, etc

Table: 1- files Chunking, Hashing, poisoning, and download Policies in P2P Content Networks

2.2 Content poisoning in P2P networks: -

Some anti P2P file sharing systems to stop the illegal file distribution tried methods such as pollution and index poisoning. In P2P sharing system have illegally shared a lots of copyrighted and malware. In order to avert the file distribution to much attention attracted by the content poisoning. In content poisoning aims to complicate users by spread by decoy or infected data files in P2P network. Its effect to the networks has not been well to do works properly. Pollution and index poisoning is the method of content poisoning to controls the P2P file sharing by the help of inserting massive type data in shared files. An effective and secure manner in copyrighted contents to paid users in P2P files sharing networks. To protect such unauthorized peers distribution proposed content poisoning methods to reduce the illegal file distribution.

2.3 Content poisoning side effect in P2P networks: -

In side effect of the content poisoning in p2P networks, its generate more traffic in networks because of in duration of downloading of the file if we can't be examine the content and download the file after that we check that, and then we observe that the desired data is not be genuine means that the content is polluted by the content pollution or index poisoning. As the result, the client fade up and give up download genuine file. This is affecting the bandwidth to be increase or wastage of the bandwidth. However some P2P systems apposed this content poisoning method. This may generate more traffic over the P2P n/w.

2.4 Impact of content poisoning: -

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.

2.5 P2P Pollution classification:-

Two major categories classified into system of file sharing systems [2]

- Content Pollution
- Index poisoning or Metadata

2.5.1 Content pollution: -

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.

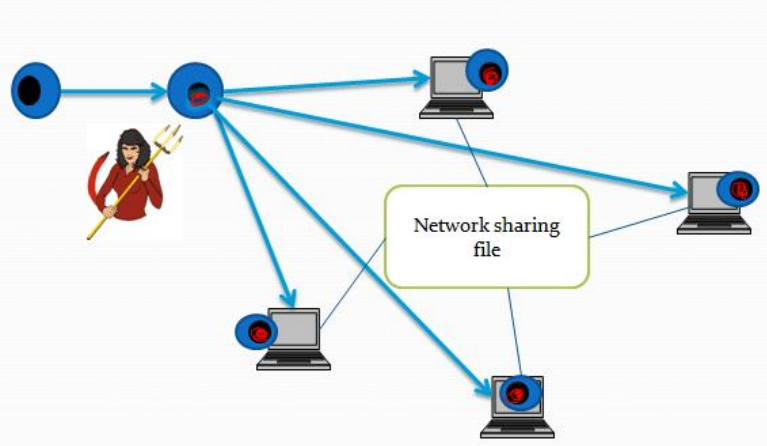


Figure 2.1: Content pollution in P2P networks

2.5.2 Index poisoning or Metadata pollution:-

In metadata an older recording take this involve often, whose copyright expired, metadata target the recent released recording from the expired copyright song with the help for changing song

title, song singer, album title into the recently release songs. When user requests target song and user will obtain the different song by mistakenly.

We also see that users obtain pollution as intentionally and unintentionally. Pollution making company intentionally create the version of pollution with the help of content pollution, metadata pollution, and index poisoning. But some it create accidently by damaging the content and then we upload in P2P sharing system. For example a user record a song from the radio and share into the P2P network and accidently at the starting or end of the song pick up the voice of the radio jockey the accidently or unintentionally obtain the pollution.

In index poisoning process of the user's aims at the index querying and in P2P networks its make stiffer to find the precise content. In index poisoning need less bandwidth and server resources as compare to content pollution. There never need to transfer nor be the response to the request. The large no. of the invalid information insert into the index obstructs. Indexes have been shared file in P2P network system. Find the location of the desired files which are searched by many users. 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).

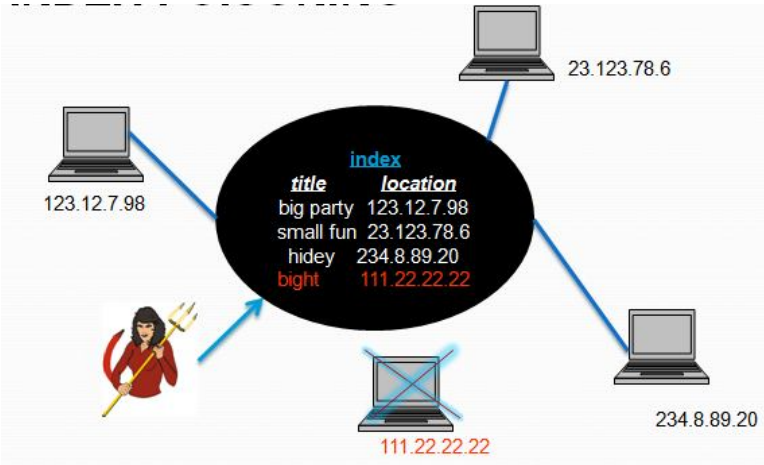


Figure 2.2: -Index poisoning in P2P Networks

In poisoning of index, can be introduced in many way, one of these method is randomly file identifies chosen which do not related to any file sharing system in any existing files.

2.6 Approaches to combat content poisoning:-

Most of the P2P systems lack effective mechanism to provide cooperation among the peers. This results in the content poisoning. To address this problem many approaches have been proposed to make P2P networks. The study of these approaches makes our mind to think about how to control the behavior of Content Poisoning in the network. The approaches proposed can be categorized into three main groups [3]:

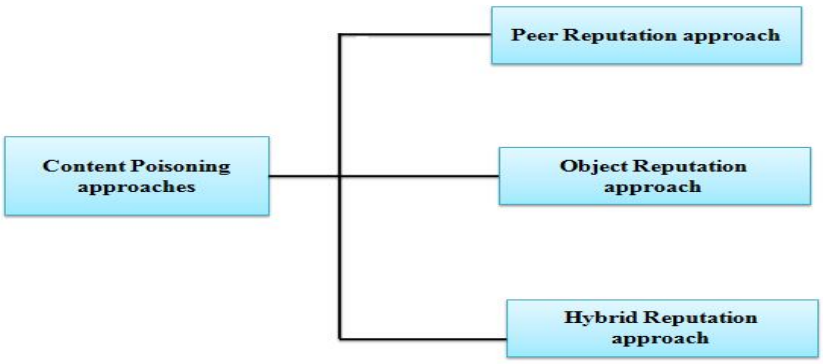


Figure 2.3: - Content Poisoning Approaches

2.6.1 Peer Reputation approach: -

P2P networks of decentralized unstructured system design of peer reputation. Peer reputation approach system guide peers that who to download a file from. We propose mechanism for peer reputation system. A dynamically updated reputation score in the P2P networks. Both mechanisms fundamentally track the reputation –score and it is used by the any peers in P2P network.

Reputation system using two scheme of computation

- Credit-only reputation computation(DCRC)
- Debit-credit reputation computation (CORC)

In first mechanism credit-only reputation (CORC) debit peer reputation score is downloading and credit serving content. In other hand the second mechanism no debit no offers, only credit peer reputation score serving content.

In P2P decentralized unstructured like Gnutella P2P system, content recovery involves a content download phase and search content phase. The desired content for search generated by a peer with suitable keyword. And it is sends to all peers. This query reply back to the peers who process the query. The request forward to the peers, if it's sharing directory has the content. The query depending on the TTL (Time-to-live) are directly connected to it. Query peer is exhausted until the TTL specified, this forwarding continues. Once all the replies receive by the querying pees. Then content download from the selected peer. At the point of download the content they use TCP connection or HTTP connection typically. For successful content re-retrieval, the type, quality, and quantity of the content each peer places in the shared directory play an important role.

Further, the bandwidth at which the actual download occurs is also an important consideration. A high bandwidth querying peer is likely to have a better experience with the system if it downloads the content from another high band-width peer. The above factors essentially differentiate the peers along the dimensions of behavior and capability. The capability of a peer depends on its processing capacity, memory, storage capacity, and bandwidth. The behavior of a peer is determined by the level of contribution offered by it for the common good of the P2P network. As the peers conduct content search and download functions, the proposed computation mechanisms map each contributing peer's behavior to form the first component of the reputation score for each contributing peer. The second component of the reputation score for each peer results from its capability (memory, power processing, band-width, and capacity storage).

Reputation based methods can be categorized into two main groups: autonomous reputation approaches and global reputation approaches. In autonomous reputation approach, peers maintain reputation history of only those peers whom they have interacted with. This approach is very easy to implement because it does not require creation of global database, security infrastructure or centralized storage to protect local reputations' integrity. Example: XRep.

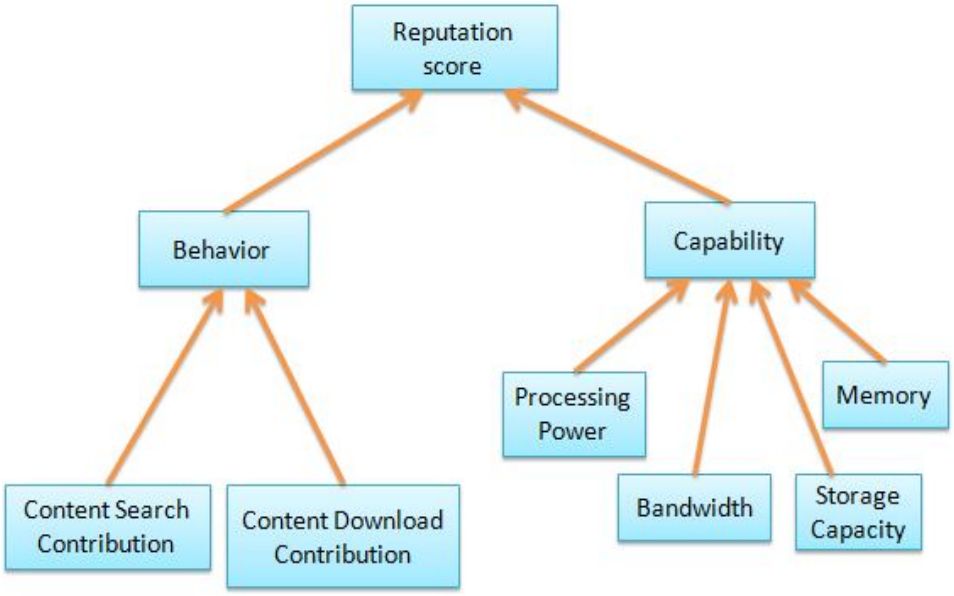


Figure 2.4: - Components of the Reputation Score

Some peers may not want to get their reputation tracked for privacy reasons. Existing designs of P2P networks do not provide peer anonymity and our goal in this paper is not to propose alternate designs for Gnutella style P2P networks. As a result, the reputation tracking presented in this paper does not address anonymity issues in such tracking. Also, the reputation system involves additional overheads to keep the most up-to-date view of each peer's reputation which some peers may not want to incur. For these reasons, enrollment in the reputation computations is self choice. Peers who choose not to enroll always maintain a default reputation score of 0.

Peers who enroll can enhance their scores by being good citizens of the P2P network. They can also save their reputation scores across sessions. Thus, a cooperative peer can maintain benefits of its participation in the system in spite of being offline for a while. In a perfect world, each peer's local software can update and store its reputation score. However, this simple mechanism could be thwarted by the peers by altering the score computations to their benefit or by tampering with the value of the stored counter. The proposed solution to prevent such occurrences is discussed. The solution utilizes a reputation computation agent (RCA) for fair periodic updates to each enrolled peer's reputation, still ensuring that the reputation points for each peer are kept locally for fast retrieval.

2.6.2 Object Reputation approach: -

For design several goals to guide that are important requirement for the successful reputation P2P mechanism.

- **Relevance: -**

  The system must require using sufficient information for credibility of the peers and the object evaluating the authenticity.

- **Distribution and Decentralization: -**

  A prior no participants should be trusted. During online operation no central computation should be required.

- **Robustness: -**

  The attacks must be robust of the system attacks by a large no. of malicious peers coordinated.

- **Isolation:** -

  The participating decision in the reputation system should be free decisions in the related activities participate such as, bandwidth contribution or online remaining and files sharing.

- **Motivation: -**

  In reputation system must have realistic incentive to participate honesty.

2.6.3 Hybrid reputation approach: -

Hybrid reputation approach works on structures of the organizational the attempts solve problems reputation approach associated with centralized and decentralized both models. In order to Organizational approaches are more and more used to build Multi Agent Systems (MAS) since they allow facing complex problems using simple abstractions. Relationships among organization members those abstractions can be concepts that structure, such as roles that agents can play and interactions that type of agents can use to connect to each others, and also bounds, such as norms that establish undesirable agent's behavior.

2.7: - Punishment:-

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 misbehavior in the first place [7].

2.8: - Summary: -

This chapter gives a brief summary of the content poisoning problem, its impact on P2P systems. We also take a look at the various approaches used to combat of the content poisoning. First of all effects are analyzed in content poisoning. The brief explanation of the various scenarios in P2P systems is given. Finally we discuss the various approaches that we have followed.

Chapter 3

Detecting Pollution and Punish Method

In this chapter, we propose a new method called detecting pollution and punish method which helps in combating content poisoning in P2P systems. In P2Pn file sharing application based on DHT. To index the file sharing files uses mechanism of double indexation for DHT. Each peer nodes associated by two level. In first associated level of DHT have keywords with files name, while the second associated level of DHT has sources with files. As we know that each node has a random ID for determining its position in the distributed hash table. When we shared any file, the given data and associate keywords with the name of its hashed generating an ID separated with MD4 function which is published in DHT. Firstly the information of the files (file name, file size, file-ID, etc) each keyword (keyword ID) is published toward the hash table. While the secondly published its own information by the peers (Peer ID, IP address, Service Port No., etc) towards the file (file-ID) of hash.[6]

3.1 Detecting Pollution in P2P Sharing Networks: -

3.1.1 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.

3.1.2 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].

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
... ...	

Table:-3 Example of consistent filenames retrieved from the responding sources for a clean file.

At the beginning of the downloading process we can obtain a file-ID linked the different file-names. After a keyword search when a file is selected for the download, firstly thanks to the DHT from the sources. Then TCP connection sends to request to initial toward each source.

Relating a clear file (table 2) the same and the others majority (30) of the sources responding desired content are clearly related. On the other side relating to a polluted file (table 3) show that the totally different filename the desired content conflicting with each other.

3.2 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:

$$K(P,Q) = \frac{|P \cap Q|}{|P \cap Q| + \alpha * |P - Q| + \beta * |Q - P|} \quad \ldots\ldots(3.1)$$

K (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 a average function of coefficient similarity for all file-name Q*i* which is regained from the sources n.

$$S(P) = 1 - \frac{\sum_{i=1}^{n} \mathrm{K(P,Q)}}{\mathrm{n}} \quad \ldots\ldots (3.2)$$

$$K(P, Q) = \frac{|P \cap Q|}{|P \cup Q|} \quad \text{…. (3.3)}$$

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 behaviors of the past content downloaded of the such peers k. in future from peers k represent the probability and maintain a list of verification. [8]

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. [1]

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 [1].

Proposed Algorithm :

1. Initialize $S(P_c^k)$ to 1.0 for all peer k.
2. For the file F from peer k.
3. With probability check $S(P_c^k)$ verify file F.
4. {
5. If
6. File is infected.
7. $S(P_c^k) \leftarrow (0.7 \sim 1.0)$.
8. Then delete file.

9. Else
10. Keep and share.
11. }
12. Otherwise
13. Keep and share.

Figure 3.1:-Pseudo code for probability check.

3.3 Summary

In this chapter, we present the details falsification and verification explaining the main approach, the proposed content poisoning levels, the detection mechanism, and the probability check. As a summary, falsification requires the server to monitor the network peers attached to it. By monitoring the behavior of its peers, a server can decide if a neighbor is acting like infected peers. If a server determines that a peer is acting as infected peers, the peer takes probability check against that neighbor to reduce the adverse effects of content poisoning. To evaluate falsification in a P2P network environment, we have conducted brief implementation tests. In Chapter 4, we present and discuss the implementation results of the proposed solution in detail.

. Experimental Results

4.1 Simulations

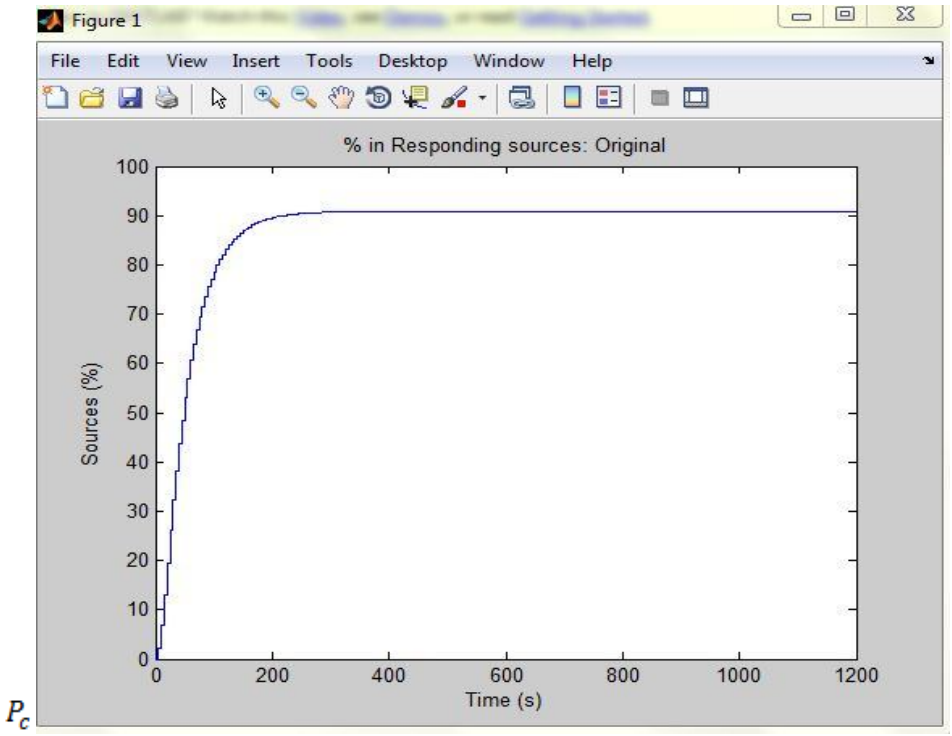
We have simulated our framework on the overlay model of Omnetpp, Inet and Oversim. We have also done small implementation of MATLAB.

4.2 Result

We have done the simulation of our framework considering a few nodes. We have observed the result by making a comparison between other approaches and our approach. The results depict that our approach is advantageous because it control the activity of content poisoning in P2P systems.

4.3 Quantification of Content Pollution in P2P Networks: -

4.3.1 Shared contents investigating: - It is an impossible task to collecting all the shared files in a P2P networks. We are trying to investigate that content pollution which is interest of the user's significative sample based. Our experiment based on the top 100 which was more downloaded contents in previous year, Bit-Torrent website indexing receives in a year more than 100 million searches. We are collecting 20 related files content from the top 100 for the each content that valued sources show the highest number, in the result of investigated 2000 files.
Previously collecting another filenames for the 2000 files, we want to know how faster file found the real sources in order duration define of the experiment. On bases of 150 sample files, the quickly increase the responding sources in the duration before
200s after that downloading they are stable that are shown in figure (9). The real detection can be performed is how faster that is show in this result.



P_c

Figure: 4.1- % of the responding sources discovered in time

4.3.2 Metric evolution: -

Doubt disposed to errors in every classification. Some unpolluted (false-negative) files are tagged as polluted (false-positive) in our case and doubt; unpolluted files can be attached as polluted (false-positive). To provide a consistent detection pollution. We have to need an appropriate fact for the corresponding metric and recognition thresholds value. By investigated file-names files tag as clean, unclassified or polluted within a web boundary which showing the connected file-names is presented in table II and table III

We tasted some Tversky forms correspondence metric and initiate that best result detection depending to the expert votes given for $\alpha = \beta=1$ which is known as Tanimoto coefficient [11] and written as :

Then to set the detection thresholds for the top contest the skilled votes: for when pollution coefficient S (P) below the 0.1~0.3 that means tabbed as a clean files, in between the 0.3~0.7 means tabbed as a unclassified files and 0.7~1.0 means tabbed as a polluted files.

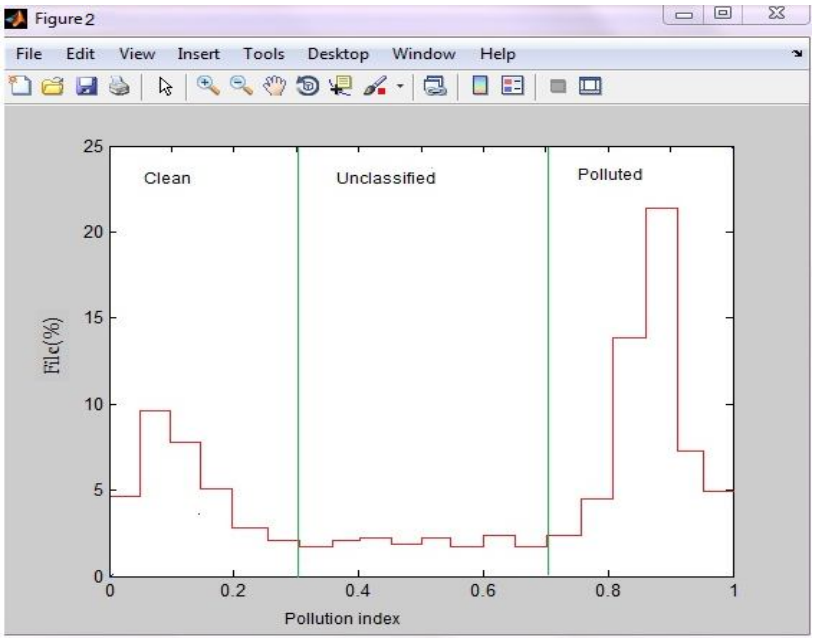


Figure: 4.2 - Distribution of files according to the pollution index

We suppose that each peers before downloading they will be verifying the content from the file if file to be infected then we have to delete from the peers but some time many uses is not skip the verification step and it shared files rest network.

4.3.4 Computerization and description of P2P Pollution:-

The index falsification to quantify, we applied to investigated our final metric on the popular files. A large number of broadcast sources split, we didn't find any response sources for 20.5% files this clearly indicates that files is the polluted from the index poisoning. Another we found that the majority of the index falsification of the pollution affecting 41.1% which are considered in P2P networks. 28.6 % are clean and 8.6 % files are unclassified.

Type	quantification (%)
No responding source (index poisoning)	20.5
Polluted (index falsification)	41.1
Clean	28.6
Unclassified	8.6

Table 4:- Quantification for the global pollution

In conclusion, we investigate for each entry in top 100 for corrupted file. We are using two types of lists keywords. One is paedophic content [6] and other is pornographic contents. Now we search for those keywords file-names 41.1% files are infected by the index falsification. And 8.8% referenced at least paedophic and rest 55.7% related to pornographic. [9]

Contents	quantification (%)
Child pornography	8,8%
Pornography	55,7%
Other	35,3%

Table 5:- Index falsification contents types

Analysis of the corrupted file of the top 100 entries, the pollution concerned of the top100 for each entries. We observe that the 25% of the most downloaded file are infected out of 20 files

Fig. 4.3 the performance shows in fraction verifications for tasted four situations vs. simulation time, and fig. 4.4 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.

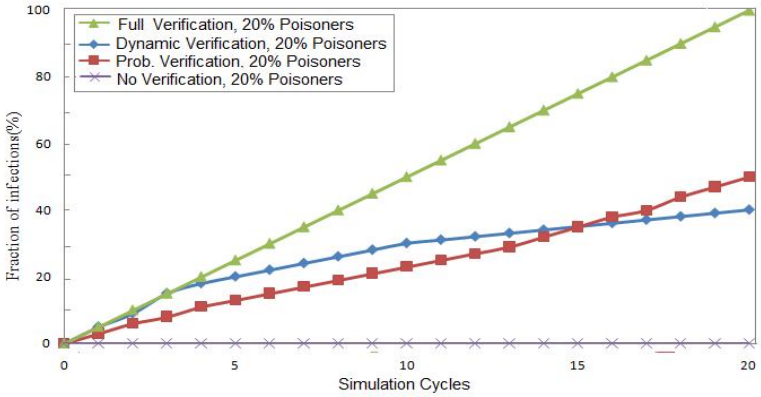


FIGURE 4.3:- fraction of the verification in the care of 20% poisoners sharing infected files

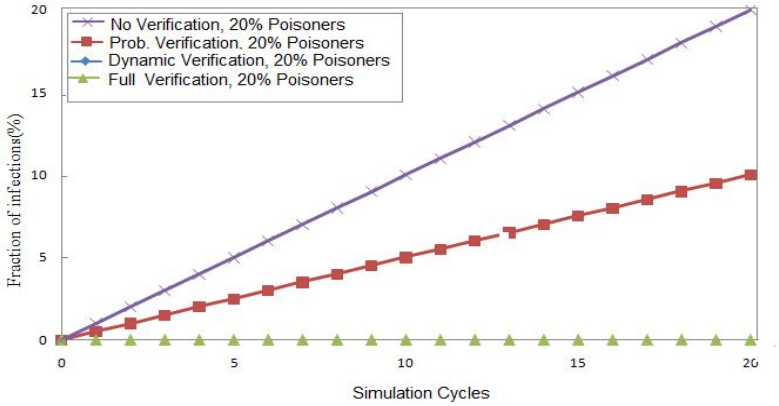


FIGURE 4.4:- fraction of the infections in the care of 20% poisoners sharing infected files

With the being said, our proposed algorithm work in the case where poisoners send some genuine file along with infected files therefore, we repeat our simulation experiments with same parameters described earlier, poisoners peers create and share 2 genuine and 8 infected files the result run on new simulated fig 4.5,4.6.

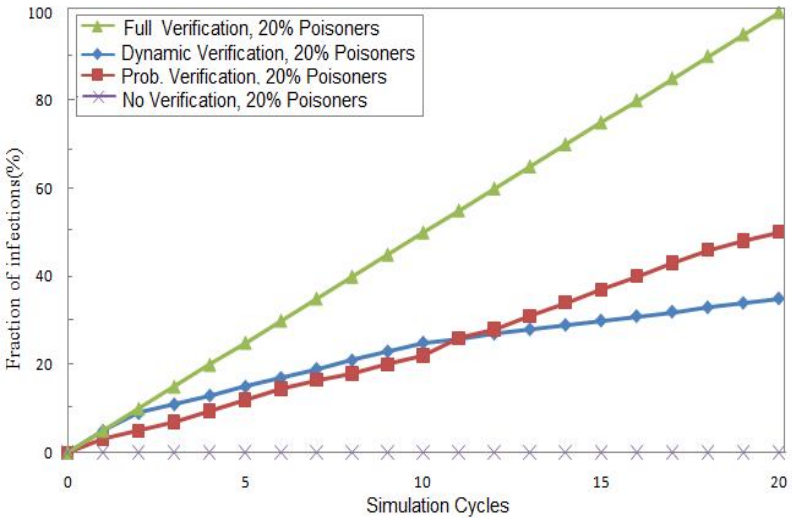


FIGURE 4.5:- fraction of the verification in the care of 20% poisoners sharing both genuine and infected files

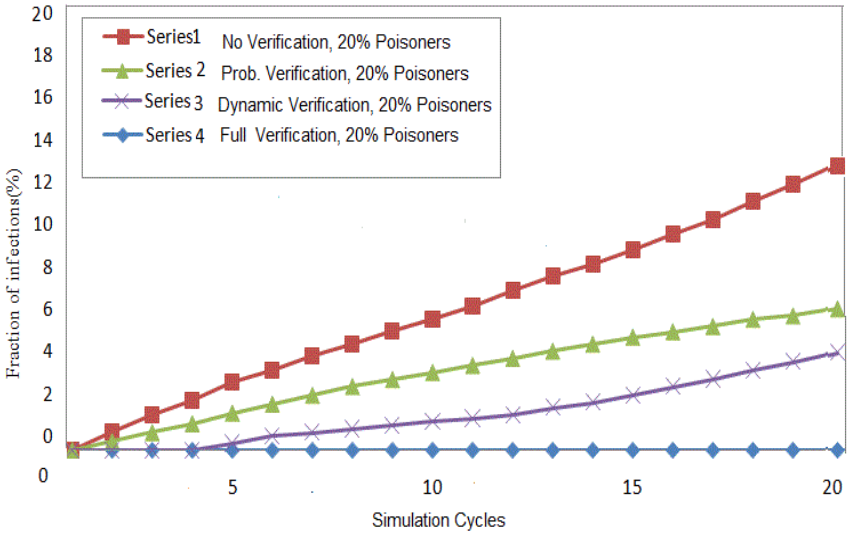


FIGURE 4.6:- fraction of the infections in the care of 20% poisoners sharing both genuine and infected files

Notice that even when the poisoners are acting as bogus peers in the system, our verification method performed quite well.

4.4 Comparison based on percentage of content poisoning

If no verification probability check is implemented in the network. Many previous researches show that almost 80% of nodes become content poisoning in the absence of verification mechanism. This increases the number of queries in the system and it can become overloaded. If the verification mechanism is able to reduce the number of content poisoning in the network, it will enhance the performance of the network on the whole.

4.5 Increment of P2P Networks efficiency

When we notice that the peers believe or act like as rogue poisoners, then our propose method of the verification are work very quiet and properly. Then the infection ratio is more the reduced up to 6% to the previous verifications (no verification and probability verification), and overhead verification is just 38% only. Apply verification algorithm is flexible and worth in monitoring and design, suppose the probability check value can be drop to prevent infection slowly, or verification overhead quickly to reduce. After the verification algorithm method apply it is more efficient because they reduce the overhead in to the networks of the P2P.

4.6 Time reduce in Networks

If the network efficiency is increase then the download or searching time also to be reduce in the networks and easy to get the files to the user. When user get clean data then traffic are also controlled and that why user are more interacted to use the network of the P2P.

Content verification reduce traffic overhead that the profit of the user and the networks to control the traffic of the network, and that decrease or reduce the time of the user uses and increase the believer of the Peers.

4.7 SUMMARY

In this chapter we give an explanation of the simulation performed to obtain the results of the research work. We explain the various results with the help of graphs. The graphs show the reason why our approach is better than the previous approaches.

Chapter 5

Conclusion and Future Works

5.1 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.

5.2 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.

14. Authority system to prevent privacy protection in Peer-to-Peer network system, S. Uvaraj1, N. Kannaiya Raja2: American Journal of Networks and Communications 2013; 2(3): 67-72 Published online June 20, 2013 (http://www.sciencepublishinggroup.com/j/ajnc) doi: 10.11648/j.ajnc.20130203.13.

15. Implementing a Distributed Peer to Peer File Sharing System using CHEWBACCA – CHord, Enhanced With Basic Algorithm Corrections and Concurrent Activation Matthew Baker, Russ Fink, David Trimm, Adam Whisman In Partial Fulfillment of the Requirements of CMSC 621, Advanced Operating Systems Fall, 2003 1.

16. Content Availability, Pollution and Poisoning in File Sharing Peer to Peer Networks Nicolas Christin S.I.M.S., UC Berkeley christin@sims.berkeley.edu Andreas S. Weigend Weigend Associates LLC andreas@weigend.com John Chuang S.I.M.S., UC Berkeley chuang@sims.berkeley.edu.

17. A Distributed and Monitoring-based Mechanism for Discouraging Free Riding in P2P Network Tian Junfeng, Yang Lidan, Li Juan Network Technology Institute Hebei University Baoding, China jftian@hbu.cn, mousekidcn1984@163.com Liu Zhongyu The fifth engineering Co., LTD 18th Bureau, China Railway China zhongyu_2003@163.com : 2009 Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns.

18. The evaluation of index poisoning in BitTorrent Jie Kong School of Computer Science and Engineering Northwestern Polytechnical University Xi'an, China jackeykong@mail.nwpu.edu.cn Wandong Cai School of Computer Science and Engineering Northwestern Polytechnical University Xi'an, China caiwd@nwpu.edu.cn Lei Wang School of Computer Science and Engineering Northwestern Polytechnical University Xi'an, China nwpu_wl@163.com : 2010 Second International Conference on Communication Software and Networks .

19. A Quick Detection of Colluders in P2P CDNs to Avoid an Illegal Leak of the Contents Ervianto Abdullah, Satoshi Fujita Graduate School of Engineering, Hiroshima University Kagamiyama 1-4-1, Higashi-Hiroshima, 739-8527 Japan Email: {ervianto,fujita}[at]se.hiroshima-u.ac.jp: 2010 First International Conference on Networking and Computing.

20. Securing Peer-to-Peer Content Sharing Service from Poisoning Attacks Ruichuan Chen1, 3, Eng Keong Lua2, Jon Crowcroft2, Wenjia Guo1, 3, Liyong Tang1,3, Zhong Chen1,3 1Institute of Software, School of EECS, Peking University, China Email: {chenrc, guowj, tly, chen}@infosec.pku.edu.cn 2 Computer Laboratory, University of Cambridge, United Kingdom Email: {eng.keong-lua, jon.crowcroft}@cl.cam.ac.uk : Eighth International Conference on Peer-to-Peer Computing (P2P'08).

3Key Laboratory of High Confidence Software Technologies, Ministry of Education, China

21. Proactive Content Poisoning To Prevent Collusive Piracy in P2P File Sharing Xiaosong Lou, Student Member IEEE and Kai Hwang, Fellow IEEE Computer Society : IEEE TRANSACTIONS ON COMPUTERS, TC -2007-09-0492R2, REVISED APRIL 8, 2008.

22. Controlling File Distribution in The Share Network Through Content Poisoning Masahiro Yoshida †, Satoshi Ohzahata □, Akihiro Nakao §, Konosuke Kawashima † †*Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei-shi, Tokyo* 50008646142@st.tuat.ac.jp k-kawa@cc.tuat.ac.jp □*The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo* ohzahata@is.uec.ac.jp §*The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo* nakao@iii.u-tokyo.ac.jp: 2010 24th IEEE International Conference on Advanced Information Networking and Applications.

23. Participation Incentive Mechanisms in Peer-to-Peer Subscription Systems S.M. Lui, Karl R. Lang, and S.H. Kwok Department of Information and Systems Management School of Business and Management The Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong SAR {imcarrie, klang, jkwok}@ust.hk : Proceedings of the 35th Hawaii International Conference on System Sciences – 2002.

24. Modeling and Analysis of P2P Content Distribution under Coordinated Attack Strategies Peiqing Zhang, Bjarne E. Helvik Centre for Quantifiable Quality of Service in Communication Systems(Q2S) Norwegian University of Science and Technology (NTNU), NO-7491, Trondheim,

Norway {peiqing.zhang, bjarne}@q2s.ntnu.no : 7th IEEE International Workshop on Digital Rights Management Impact on Consumer Communications (DRM 2011).

25. A Quick Detection of Colluders in P2P CDNs to Avoid an Illegal Leak of the Contents Ervianto Abdullah, Satoshi Fujita Graduate School of Engineering, Hiroshima University Kagamiyama 1-4-1, Higashi-Hiroshima, 739-8527 Japan Email: {ervianto,fujita}[at]se.hiroshima-u.ac.jp : 2010 First International Conference on Networking and Computing.

List of Publication

[1]. Content Poisoning in Peer to Peer Networks.

(http://www.researchpublication.com/journal/IJCSIT/Issue-2-April-2014-June-2014/0)

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

CURRICULUM VITAE

PRATIBHA SINGH

117 K/14-B R.S. Puram Sarvodaya Nagar, Kanpur

Pin Code-208025.

Email-pratibha1606@gmail.com,

Contact No-08896190502, 09415441073.

Education

- Pursuing M.Tech in Computer Science and Engineering (Computer Network) from Babu Banarasi Das University Lucknow with CGPA of 7.00.
- B.Tech in Computer Science and Engineering from MAHARANA PRATAP ENDINEARING COLLEGE (Uttar Pradesh Technical University,Lucknow) with 67% in 2011.
- Intermediate from B.S.S Inter College 78.2% in 2007.
- Secondary School from Kanpur Vidya Mandir, Kanpur with 57% in 2004.

Skills

Environment: Windows9X//2000, XP, 2007
Programming Language: C, C++, Java, Visual Basic 6.0, android, php, .net
Web Technology: HTML**,** jsp, servlet
MS-Office (MS Word, MS Excel & MS PowerPoint)

Achievements

Paper published : Content Poisoning in Peer to Peer Networks(http://www.researchpublication.com/journal/IJCSIT/Issue-2-April-2014-June-2014/0