Self Certification Based BloomCast File Retrieval In Unstructured P2P Networks

Nanitha Mathew, V.Gowri

Abstract: Peer-to-peer (P2P) networks are exposed to peers who fake, advertise malign code, or peers who won't contribute. The classic security techniques like client-server networks are not enough for Peer-to-Peer networks because of its centralized nature. BloomCast is a full-text revival scheme which is used in unstructured P2P networks. It duplicates Bloom Filters rather than duplicating the documents among th e network. BloomCast greatly minimizes the communication cost needed for duplication. But the main disadvantage is that it is vulnerable to malicious peers. In the proposed method self certification technique is used. Self-certification removes the need of a centralized system for sending the identities. The Self Certification mechanism helps to identify the malicious content and makes the search more effective.

Index Terms: BloomCast, File Searching, P2P, Self Certification

————————————◆————————————

1 INTRODUCTION

The peer-to-Peer network is a distributed architecture which divides the load among each and every peer. The Peers are equal participants in the application. The peer-to-peer network of nodes is hence formed. All the nodes are equivalent (Peers). Data could be present at any node on the network. The network is quite flexible. With the popularity of Napster and Gnutella which are peer-to-peer file sharing applications, are used by many people for searching the data needed. The Peer-to-Peer networks can be categorized into centralized and decentralized depending on the existence of a central server. Both the resource exploration and download are divided in the decentralized architecture. It results in increased robustness and scalability. The decentralized P2P networks can be again classified as structured and unstructured networks. In structured networks, there are certain rules on placing the contents. In unstructured P2P networks, there is no limitation on the placement of contents and is not related to the network or the topologies. The Unstructured P2P networks will do better in environments that are dynamic. In Peer-to-Peer a network of peers provides a search service collaboratively Object search is very important in several of the Peer-to-Peer applications. Peer-to-Peer networks can be best used for full-text revival. This is because the nodes that have the documents can be able to handle the query assessment operations Gnutella is a very popular Peer-to-Peer search protocol. Specifically, because Gnutella networks are unstructured and the peers participating in networks connect to one another randomly, peers search objects in the networks with message flooding. Various file searching techniques are flooding, replication strategies, superpeer architecture, interest locality, DHT based. In flooding every incoming message is forwarded to every outgoing line but the one from which it arrived on. In Replication strategy the items and queries are replicated properly across the network. The search success rate is greatly improved and at the same time can avoid flooding the unstructured Peer-to-Peer networks With the superpeer architecture the peers with added memory, processing capacity and network connection range provide distributed search services for discovering the resources. Thus, the peers with definite resources would not become a hindrance in the searching network. In interest locality, if a particular peer has certain information which a user is needed, then there is a chance that the same peer will be having other contents that the user has interest. With the DHT based searching, the request for documents is made, using a key. The keyword queries are mapped to the particular routing keys. This is done by mapping the node in the DHT that will store a list of documents to each keyword. your paper.

2 RELATED WORK

Hanhua Chen, Hai Jin, Xucheng Luo, Yunhao Liu, Tao Gu, Kaiji Chen and Lionel M. Ni proposed a new technique called BloomCast[1]. It helps in reviving the full-text. The BloomCast combines the distributed hash table and the unstructured Peer-to-Peer networks. Bloom Filter is duplicated instead of duplicating the raw documents. A bloom filter is a data structure which is efficient. By duplicating the encrypted sets using the Bloom filters, both the communication as well as the storage costs is greatly decreased. This also supports the full-text multikeyword searching. Two conditions need to be satisfied for the working of BloomCast technique. They are: 1) The query and document duplicates are frequently divided among the P2P networks 2) Each and every peer is aware of the size of the network.. In the Bloomcast network theBloomCast peer has a local database where it stores the documents. A bootstrap node stores the incomplete list of BloomCast nodes which is present in the system. The bootstrap node is first contacted by the bloomCast node for joining.. Structured peers form a global DHT which are a small fraction of peers with good connectivity by the bootstrap peers. A normal peer looks it local database. For exhibiting the document in a compact form it generates Bloom Filter of documents. The state information of normal peers is stored by the DHT in BloomCast nodes. For replication a normal peer first makes use of the DHT nodes which are connected. This is used for sampling an adequate amount of peers and inserts the replica in the form of a Bloom Filter to the chosen random nodes. Andrei Broder and Michael Mitzenmacher proposed the concept of applications of Bloom Filters [2]. A Bloom filter is a simple and efficient data structure. This is used for supporting the membership queries. A bloom filte alsor shows the elements which are not present in the set called false positive. The Bloom filter symbolizes a set or a list of items. A Bloom filter exhibit a set that can reduce space, but at the cost of possibility of bringing false positives. If false positives are not a problem, then the Bloom filter provides improved performance. The filter starts as an array where all the bits are set to 0s. Each item in the set is hashed k times. A bit location is returned on each hash and those bits are set to 1. The concept of self certification [3][4] is used for assuring secure and appropriate availability of the reputation information of a peer. In unstructured P2P networks there is a possibility of malicious codes and false transactions. It generates false identities in order to perform false transactions with other identities. Whether a peer is malign or not can be determined by the peers reputation. The transaction is aborted on detecting malicious content. The identity is associated to the reputation of a peer. Self-certification is used thereby generating the identity certificates. Each peer has got their own certificate authority. The certificate authority issues the identity certificate and digital signature to the peer.

3METHODOLOGY

In this section a detailed architecture is presented. The architecture is as shown in fig 1. The Self Certification based bloom cast file retrieval has got three modules.

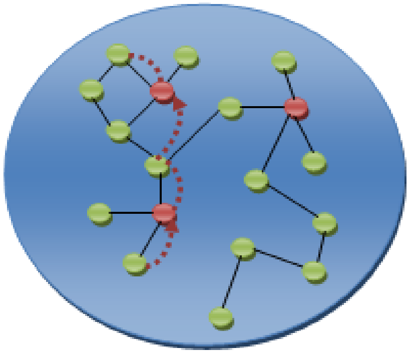


Figure 1:- Architecture

A. Peer Creation

Peer-to-peer (P2P) networking is a distributed architecture that divides the tasks between each and every peers. All the peers are equal. The peer handles the request and response queries. The nodes are created in the network for reviving the full text. Each node sends the documents frequently and at random in the unstructured Peer-to-Peer networks. The nodes are created in the unstructured P2P networks thereby minimizing the communication as well as the storage cost.

B. Content Searching and Bloomcast

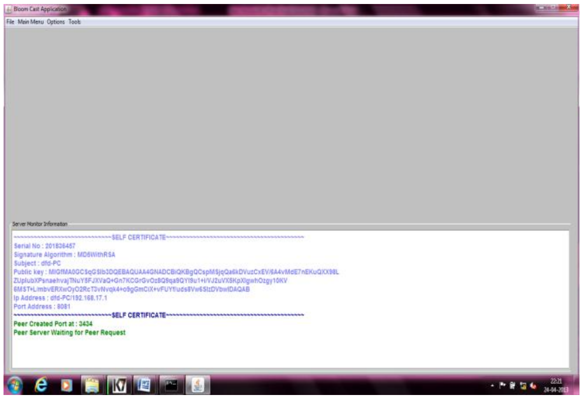
In this module the requester will send the request and the corresponding service is received from the provider. Bloomcast is produced based on the network size, node subset sampling, and replication protocol and query evaluation. The request is made to the Bloomcast node. The Bloomcast node determines whether to go to the backend or not. In a Bloomcast network there are two types of nodes, BloomCast nodes and normal nodes. Instead of replicating the whole files we are replicating the path information only in the Bloomcast nodes. If a peer wants to add to or leave the network, a request has to be made to the Bloomcast node. In the case of leaving, Bloomcast node checks whether the data and path has been replicated. If so then only it allows to leave that peer. The Bloomcast node is having a hash table. There it saves the path information that is replicated.

C. Self Certification

The Self Certification mechanism helps to identify the malicious content and makes the search more effective. In self certification each peer is having their own certificate authority The certificate authority issues the identity certificate and digital signature to the peer. The self certification is used for assuring secure and appropriate availability of the reputation information of a peer. The certification uses the concept of RSA [5] and DSS, where the algorithm generates both the private key as well as the public key. The sender encrypts the information or file using its own private key and receiver's public key. The receiver then decrypts the file using its private key and sender's public key. If certificate mismatch, then the content has been modified.

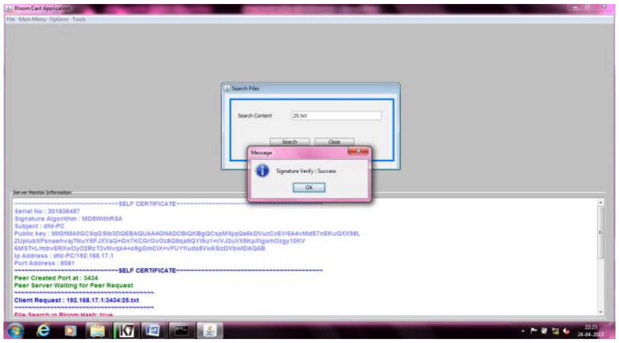
4 EXPERIMENTAL RESULTS

P2P networks divides the task within the peers. All the peers are equal. First the peer process is created at the specified port as shown in fig [2]. The fig [3] shows the search using the Bloomcast technique and returning the requested file with self certification. The request is made to the Bloomcast node. The Bloomcast decides whether to go to the back end. The file is requested and Bloomcast node returns the requested file if the file is present. If file is not there then a message is displayed. The path information is stored in the Bloomcast nodes hash table. The peer generates the public key and private key using RSA algorithm. The sender encrypts the information or file using its own private key and receiver's public key. Similarly the receiver decrypts the file using its private key and sender's public key. The MD5 is mainly used to provide Digital Signature. The larger messages are compacted before encrypting. If a hacker modifies the content then the certificate mismatch occurs and then a message is displayed saying the content has been modified. The modification details will be shown as in fig [4].



```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SELF CERTIFICATE~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Serial No : 201836457
Signature Algorithm : MD5WithRSA
Subject : dfd-PC
Public key : MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQCspMSjqQa6kDVuzCxEV/6A4vMdE7nEKuQXX98L
ZUplubXPsnaehvajTNuY5FJXVaQ+Gn7KCGrGvOz8Q9qa9QYI9u1+i/VJ2uVX5KpXIgwhOzgy10KV
6MST+L/mbvERXwOyO2RcT3vNvqk4+o9gGmCiX+vFUYf/uds8Vw6SIzDVbwIDAQAB
Ip Address : dfd-PC/192.168.17.1
Port Address : 8081
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SELF CERTIFICATE~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Peer Created Port at : 3434
Peer Server Waiting for Peer Request
```

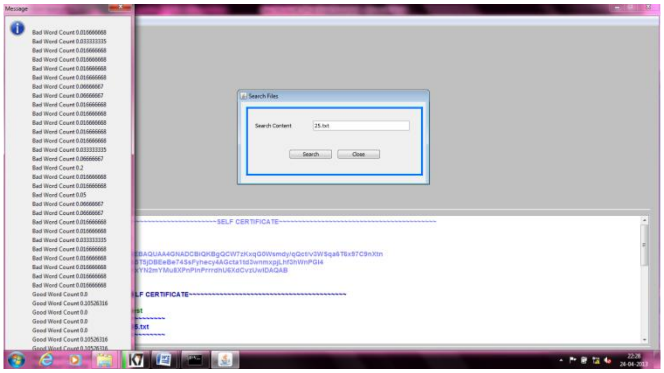
Fig 2: Peer process is created at the specified port



Bloom Cast Application
File Main Menu Options Tools
Search Files
Search Content
25.txt
Search
Close
Message
Signature Verify : Success
OK
Server Monitor Information

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SELF CERTIFICATE~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Serial No : 201836457
Signature Algorithm : MD5WithRSA
Subject : dfd-PC
Public key : MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQCspMSjqQa6kDVuzCxEV/6A4vMdE7nEKuQXX98L
ZUplubXPsnaehvajTNuY5FJXVaQ+Gn7KCGrGvOz8Q9qa9QYI9u1+i/VJ2uVX5KpXIgwhOzgy10KV
6MST+L/mbvERXwOyO2RcT3vNvqk4+o9gGmCiX+vFUYf/uds8Vw6SIzDVbwIDAQAB
Ip Address : dfd-PC/192.168.17.1
Port Address : 8081
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SELF CERTIFICATE~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Peer Created Port at : 3434
Peer Server Waiting for Peer Request
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Client Request : 192.168.17.1:3434:25.txt
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
File Search in Bloom Hash: true
```

Fig 3: content is searched using the Bloomcast technique and is send to the receiver using self certification



```
Bad Word Count 0.016666668
Bad Word Count 0.033333335
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.06666667
Bad Word Count 0.06666667
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.033333335
Bad Word Count 0.06666667
Bad Word Count 0.2
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.05
Bad Word Count 0.06666667
Bad Word Count 0.06666667
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.033333335
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Bad Word Count 0.016666668
Good Word Count 0.0
Good Word Count 0.10526316
Good Word Count 0.0
Good Word Count 0.0
Good Word Count 0.0
Good Word Count 0.10526316
Good Word Count 0.10526316
```

Search Content: 25.txt

```
~~~~~~~~~~~~~~~~~~~~~~~SELF CERTIFICATE~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

EBAQUAA4GNADCBiQKBgQCW7zKxqG0Wsmdy/qQctrv3WSqa6T6x97C9nXtn
5T5jDBEeBe745sFyhecy4AGcta1td3wnmxpjLhf3hWnPGI4
XYN2mYMu8XPnPInPrrrdhU6XdCvzUwIDAQAB

LF CERTIFICATE~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
st
~~~~~~~~
5.txt
~~~~~~~~
```

Fig 4: modified data details

5 CONCLUSION

BloomCast, a full-text revival technique, is used in the unstructured type of P2P networks. The communication cost needed for the full text search is minimized. Moreover in the place of documents the bloom filters are duplicated among the network. This reduces the communication cost needed for the duplication. But the BloomCast has the disadvantage that the peers are exposed to malicious content. A new scheme called Self certification is proposed which provide security improvement to the BloomCast file searching. With Self certification all the peers in the P2P network are identified by identity certificates. Each peer has got their own certificate authority. The certificate authority issues the identity certificate and digital signature to the peer.

REFERENCES

[1] Hanhua Chen, Hai Jin Senior, Xucheng Luo, Yunhao Liu, Tao Gu, Kaiji Chen, and Lionel M. Ni (2012) ‘BloomCast: Efficient and Effective Full-Text Retrieval in Unstructured P2P Networks’ IEEE transactions on parallel and distributed systems,Vol. 23 No. 2

[2] Andrei Broder and Michael Mitzenmacher ‘Network applications of bloomfilter: A survey’ Internet Mathematics Vol. 1 No. 4 pp. 485-509

[3] Arulkumar C V, Jeyakumar K, Malarmathi M. and Shanmugapriya.T. (2012) ‘secure communication in unstructured P2P networks based on reputation management and self certification’ International Journal of Computer Applications Vol. 44 No.15

[4] Prashant Dewan and Partha Dasgupta (2010) ‘P2P reputation management using distributed identities and Decentralized recommendation chains’ IEEE transactions on knowledge and data engineering, Vol. 22 NO. 21

[5] Rania Salah El-Sayed and Mohammad Ali Gomaa (2008) ‘An Efficient Signature System using Optimized RSA Algorithm’ IJCSNS International Journal of Computer Science and Network Security, Vol. .8 No.12.