Efficient Data Communication and Storage using Synchronized Servers with Load Distribution Facility

Prof. Aparna Kalaskar[1], Mohit Patil[2], Smit Mahajan[3], Neha Goundadkar[4], Amit Raj[5]

Assistant Professor, Department of Computer Engineering, Sinhgad College of Engineering, Pune, India [1]

Student, Department of Computer Engineering, Sinhgad College of Engineering, Pune, India [2,3,4,5]

Abstract: In the early 2000's, when data used to be of simple structure and small size, the client- server architecture handled the sharing and storage of data pretty efficiently. But as the technology progressed, the data became more and more complex and huge in size which led to load on servers for storing and sharing the data. Here, the P2P network architecture comes into picture where the peers could communicate and share the data without any centralized server. The basic idea behind the proposal of this system is the combined use of segmented file transfer on P2P network. This makes it possible to download a single file from multiple peers at the same time. Moreover, it also includes a synchronized server module for efficient storage and updating data more efficiently on the servers.

This paper presents a system in which data transfer and storage system uses file downloading based on peer-to-peer network architecture. It's the software developed for sharing/downloading huge data files and storing important data efficiently. It's very convenient to share data using this system as the data is spread across the peer-to-peer network. The synchronized servers provide reliable data storage for any file format. All the peers present in the network communicate to other peers using the information present in the tracker. Tracker is responsible for maintaining the peers in the whole network.

Keywords: Client-Server, Peer-to-Peer, Synchronized servers, Load distribution.

I. INTRODUCTION

The past few decades have experienced a huge paradigm shift from traditional client-server systems to peer to peer (P2P) network. In a client-server-based system, greater bandwidth is consumed as the demand for a particular file increase. Whereas in a P2P system, as the demands for a file increases, the nodes seeding that file also rise thus decreasing the delivery cost per file distributed. A P2P network is faster, robust and reduces the load on a single server.

Segmented Data Transfer is a technique of breaking files into chunks, downloading these chunks of the file simultaneously from a source and aggregating them on the client who requested the download. A file downloaded by using the concept of segmented file transfer is observed to have much lesser download time. The major advantage of using the concept of segmented file transfer (both upload and download) files over the internet is that the user will be able to utilize its internet connection to the maximum as the amount of data that is wasted will be reduced as well as the amount of time needed to transfer the file will also reduce without compromising on the security of the data.

BitTorrent, one of the most popular P2P technology, makes the distribution of files much easier by consuming less bandwidth from the publisher. A meta info file (called the torrent) file is created which contains information like file name, file size, hashing details along with the URL of the tracker. The download is initiated by opening this torrent file in the BitTorrent client (a free application). The system consists of a tracker which maintains the details of all the peers involved and helps to find peers. The seed (a user who initially has the file to be downloaded) uploads the entire file. Once an entire copy is distributed amongst the other downloading users, the seeder can stop uploading the file while allowing other people to download. As discussed previously, while making the torrent file, the original file is divided into smaller pieces and the hash codes of these pieces are included in the torrent file.

II. RELATED WORK

Several file sharing recommendation ideas based on user requirement have been proposed in the last decade.

A. Domain:
   The overall idea of client-server systems to Peer to Peer (P2P) network includes data transfer techniques of breaking files into chunks, downloading chunks, synchronization of servers and load distribution. Hence the idea of client-server systems to peer to peer (P2P) network comes in **Computer Network**.

B. Related Study:

After doing a research on the papers on peer-to-peer file sharing efficient data communication and storage using synchronized with load distribution.

In paper[7] from year 2020 of IEEE methodology used to increase the content availability, accelerating the download process. An improved applicable network coding scheme is referred as Super Generation Network Coding (SGNC). This is designed for file sizes ranging from 320MB to 5GB, and utilize three-piece scheduling policies namely: random, local RF, and global RF. Due to this paper pros we get are SGNC maximizes the generation size so that it is as close as possible to the optimal size without adding computational overhead and P2P content distribution systems is efficient in terms of content availability, download time, overhead, and decidability. We also get cons from this paper are Dense Network Coding (DNC) has been proposed theoretically considered as an optimal solution & have huge computational overhead, it is not viable for real- world systems. This is not for the overhead of very large files, i.e. more than 10GB's files.

In paper[8] from year 2020, in this paper of methodology used is an efficient direct and indirect file transfer protocol (C2CFTP) that transfers files between clients in a client-server system. Here the file transfer performance is be improved by selecting and using one of the proposed direct and indirect C2CFTPs depending on the situation when transferring a bundle of files. Pros of this paper are this can omit unnecessary file I/O overhead by relaying the file to the receiving client instead of storing the file at the server for indirect transmission. For direct transmission, instead of connecting a data channel every time a file is transmitted, the channel connection overhead is reduced. Cons is the file transfer is required within a predetermined time.

In paper[3] from year 2019, in this paper methodology discussed are data synchronization techniques are as follows Snapshot Replication, Merge Replication and Transactional Replication. Pros of this paper is Transaction Replication gives efficient data synchronization system and Cons of this paper is no load distribution is performed on synchronized servers.

In paper[5] from year 2019, in this paper methodology are architecture peer to peer communication network is used for distributed data transfer and it involves download different portion of file from multiple client. Pros of this paper are use of distributed data transfer results in faster data sharing as compared to centralized system, high bandwidth utilization and load balancing is performed by using custom scoring logic. Cons of this paper is less secure as there is no server to check the data over the network.

In paper[2] from year 2019, in this paper methodology used in this system the the file to be shared is divided into equal parts and these parts are then shared to same recipient via max flow paths. Pros of this paper is less time is required due to division of file in equal parts. Cons of this paper is only 1 sender can send data to n receivers at a time.

C. System Overview:

This system would be built for sharing/download huge data files and storing important data efficiency. It's very convenient to share data using this system as the data is spread across the peer- to-peer network. The synchronized servers provide reliable data storage for any file format. In this system tracker is responsible for maintain the peers in whole network. This classification model would then be implemented to share data files and storing using synchronized servers.
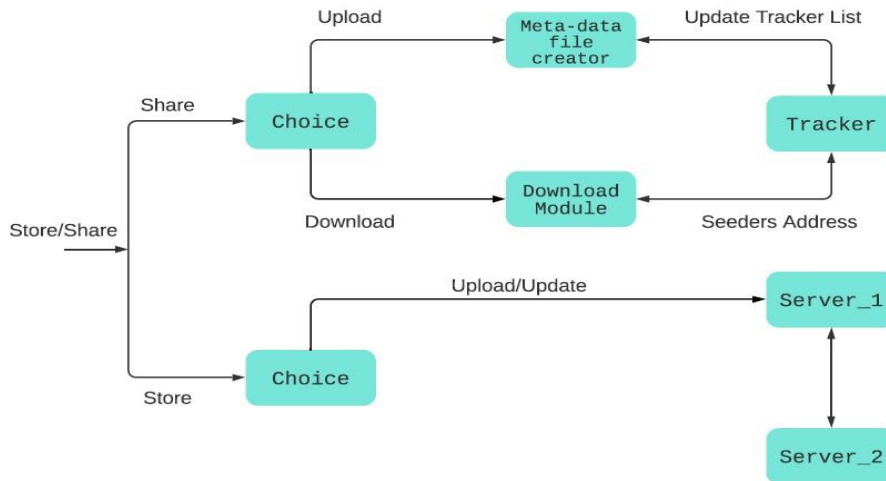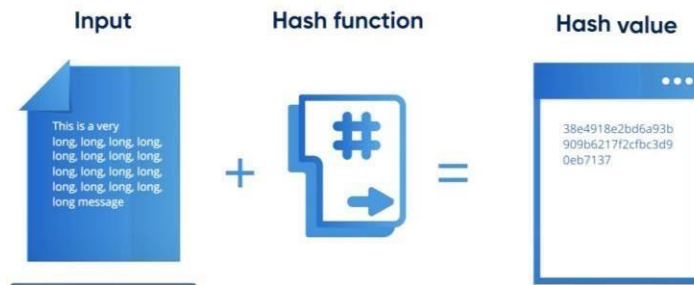
Fig. 1: System Architecture Diagram

III. MATHEMATICAL MODEL

Hashing: - It is the process of converting an input of any length into an array of fixed size stringtext or number using a mathematical function.

SHA-256: - It is one of the successor hash functions to SHA-1 (collectively referred to as SHA-2), and is one of the strongest hash functions available. It is like the fingerprints of the data. Even if only one symbol is changed the algorithm will produce different hash value. SHA256 algorithm generates an almost-unique, fixed size 256-bit(32-byte) hash.

Mathematical Function-

1. SHR 32-   Shift right input bit by 32 bits.

2. ROTR-   perform Circular right shift on input bit

3. XOR-   Perform XOR on two binary number

4. ADD-   Add two binary number

5. Lowercase Sigma zero- Perform ROTR (7 bits) on input -> ROTR (18 bits) ->SHR (3 bits)

6. Lowercase Sigma one- Perform ROTR (17 bits) on input -> ROTR (19 bits) ->SHR (10 bits)

7. Uppercase Sigma zero- Perform ROTR (2 bits) on input -> ROTR (13 bits) ->SHR (22 bits)

8. Uppercase Sigma zero- Perform ROTR (6 bits) on input -> ROTR (11 bits) ->SHR (25 bits)

9. Choice(ch)- Uses first input to decide whether to take input from first or second.

10. Majority ()- It takes three input binary string and consider majority of three bitsas input.

11. Constant- calculate using cube root of prime number (3√prime number).

$$T1= \textstyle\sum 1(e) + ch\ (e,\ f,\ g) + h + K+WT2=\textstyle\sum 0(a) + Majority\ (a,\ b,\ c)$$

Here a, b, c, d, e and f are working variables.

IV. CONCLUSION

In this paper there is implementation of distributed data sharing using P2P for the fast sharing of the huge file and storing data efficiency. It uses better utilization of allocated resources. For updating files, it required less amount of time because it uses peer to peer. Implementation of synchronized servers gives more reliability of data storage and in this system ensuring auto recovery of data.

REFERENCES

[1]   Peer Assisted Parallel Downloading System International Conference on Communication and Signal Processing, April 4-6, 2019, India.
[2]   Minimizing One-to-Many File Transfer Times using Multipath-Multicast with Reed-Solomon Coding 2019 31st International Tele traffic Congress (ITC 31).
[3]   Research on Data Synchronization of Cement Intelligent Control Cloud Platform
[4]   Generalized Secure Hash Algorithm: SHA-X
[5]   Peer to Peer file sharing
[6]   A High-Performance Parallel Computation Hardware Architecture in ASIC of SHA-256 Hash International Conference on Advanced Communications Technology (ICACT)
[7]   Super Generation Network Coding for Peer-to-Peer Content Distribution Networks IEEE-2020
[8]   Direct and Indirect File Transfer Protocols Between Clients in Client-Server Architecture