

INTERPLANETARY FILE SYSTEM

An analysis on Future File System

¹Mr. Atul Kumar Verma, ²Mr. Arpit Garg

¹Assistant Professor, ²C.E. Student

Dept. of Computer Engineering, Poornima College of Engineering,
Jaipur, India

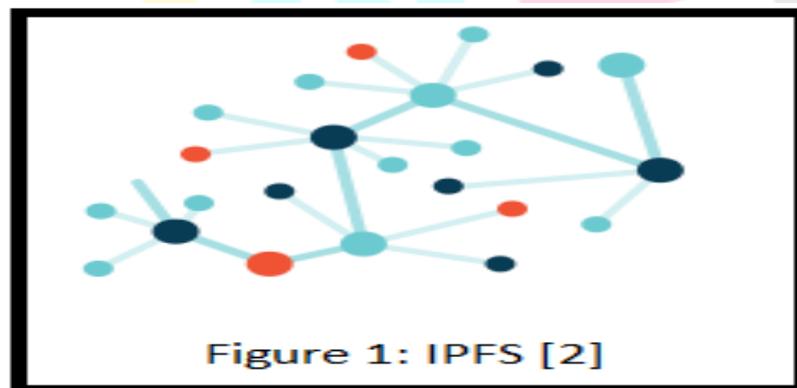
Abstract- *InterPlanetary File System (IPFS) is a peer to peer to the protocol using decentralized methods. It is hypermedia distribution protocol. Hypermedia is the advancement of hypertext. The information stored can be accessed based on its content, all the contents in this are associated with the hash key. IPFS is used to store permanent data and remove the duplicate data. IPFS has no single point of failure and prevents the data from various attacks which are frequent in HTTP like DDoS etc. All the files in IPFS are available to the world and can be easily and quickly access through the hash key every file is associated with it. It is an open-source project which focuses to make the internet faster, safer and open to all. It is a block storage model which connect all the computing devices to the same file system. FUSE (Filesystem in Userspace) can be used to access IPFS rather than using traditional HTTP.*

Index Terms--*Hypermedia, DDoS, Hash, FUSE, Open-Source(keywords)*

I. INTRODUCTION

InterPlanetary File System (IPFS) is emerging area of technology that will be soon going to occupy the market. It is a protocol that works on peer to peer technology which is decentralized and distributed. It is a block technology that provides a mechanism to store and share permanent files. It is an alternative to central storage system often termed as servers. Servers are having many drawbacks due to increase in their usage day by day. Now, to overcome all the issues related to storing and sharing of the file, there is a need for improved and advanced protocols which can change the traditional centralized system and make the systems more beneficial in terms of safety, speed, accuracy and much more. It is based on content rather than storage. It uses the concept of content-addressable storage (CAS). According to CAS the information stored that is going to be retrieved is based on its content not depend on its storage. It is a hypermedia distribution protocol. Hypermedia is the advance version of hypertext which provides the strong base for nonlinear information like audio, video or any graphical based information including non-graphical information and hyperlinks. All the information stored can be accessed using the concept of hash key. Every file stored in the system has a unique key associated with it called hash key which can be used to directly access that file from anywhere in the world. It contains files, directories and mountable file system which can be accessed through FUSE (Filesystem in Userspace) ^[1]. It provides the full authenticity to the users as it uses cryptographic hash content addressing mechanism. It fully blocks decentralized system so there is no central point of failure in the system.

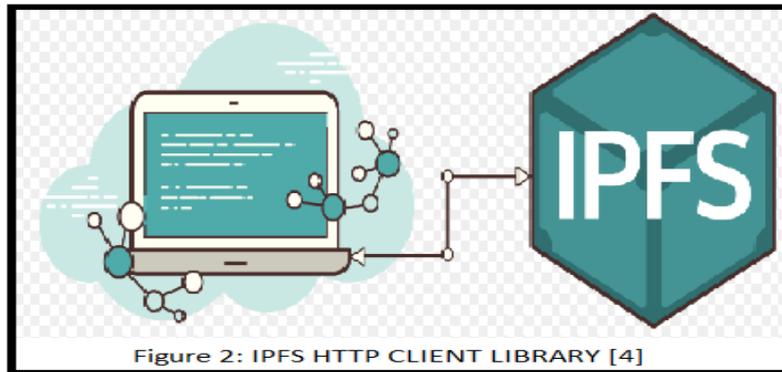
HTTP is the most successful protocol used for today's web but there are certain drawbacks of HTTP. The main disadvantage is safety HTTP is very much prone to cyber-attacks like DDoS attacks. IPFS protocol enhances the HTTP with add-on new functionality that does not degrade the user's experience. This can be used to view the documents online with file access that can be done through HTTP specified path and extensions can easily learn to use the URI schemes directly. HTTP allows moving small files for cheap, even for small organizations with a lot of traffics. In present scenario need of the organization's increases which needs to deal with the datasets which are in extremely large volumes. These datasets are not able to be managed through HTTP so there is a need IPS which can easily manage the datasets in an economical way. Bitstamp ledger is present which contains the record of all the transfers between the nodes. It allows the nodes to keep the history track and prevent from any malicious activity.



II. IPFS OVER HTTP

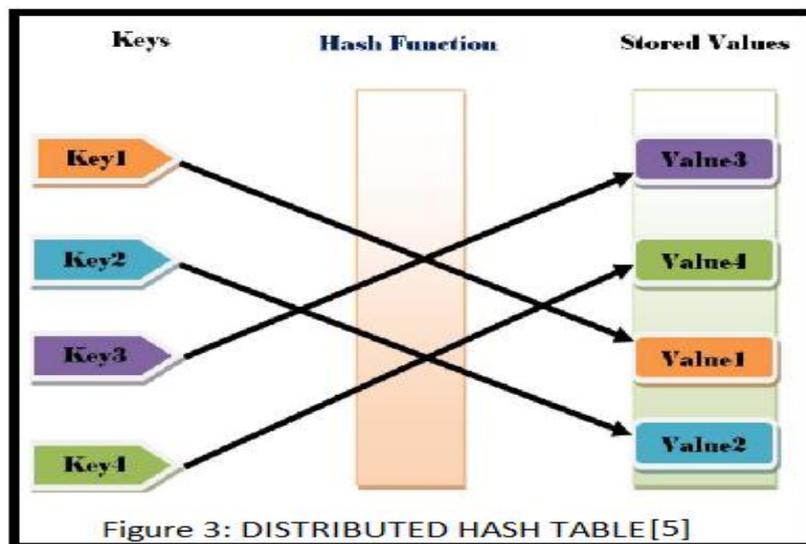
HTTP is used over a long time because of it can move small files at very low cost. Now a new technology is in demand which can deal with larger datasets like petabytes and that too be at affordable rates. This problem is solved by IPFS which gives the power to deal with large datasets and in the most economical way. There is a requirement of hosting and using the large volume of datasets which is successfully fulfilled by IPFS. Technology is increasing its level day by day and to deal with that large computing powers are required. With the increasing computing powers, there is an increase in datasets. Computing services needed a large dataset to perform their computations in an organization so IPFS make it easy and affordable with full authenticity to use within an organization.

Media Streaming platforms are increasing every minute and there are requirements for large servers and centralized system to maintain. Image quality is increasing and market demands high-definition on the go which requires more resources which increases the cost. This can be minimized by using distributed and decentralized properties of IPFS which gives high-definition on-demand media services on the go. Large data are stored on a server and these servers are located at different locations and all the data present on the servers must be linked with one another to maximize their usage, there must be a chance of errors or attacks on these large volumes of data which can be a trouble for any organization. By using IPFS one can easily perform linking and visioning of large datasets with full safety and no failure. There is a risk of disappearance of important files and documents using traditional methods but using this technology there is no chance of any failure or disappearance of any files or documents. This technology provides a large amount of data to everyone everywhere with the minimum resources required. Every file in IPFS can be found by human readable format using a decentralized naming system called Inter-Planetary Naming System (IPNS). IPNS allow adding an extra feature to ease the use of users. It allows storing a reference of an IPFS hash under the hash of your public key^[3].



III. DISTRIBUTED HASH TABLE

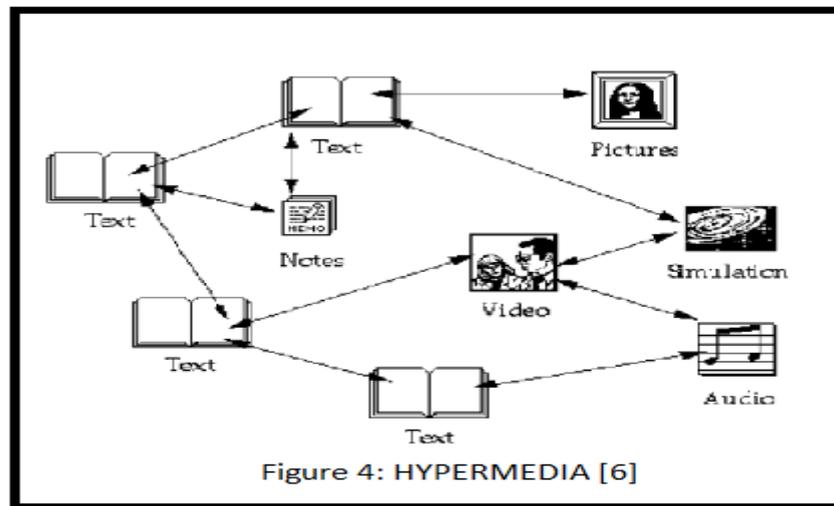
The distributed hash table is a decentralized and distributed way to store metadata about peer to peer networks. It uses the same concept as in a hash table that is look up the value using the key except that data are distributed on any number of nodes. All the nodes in the system are interconnected which make insertion and retrieval possible among all the nodes. Distributed hash tables increase the security of data and providing the method to access the data at a rapid rate without any failure. There are different types of distributed hash table present like KADEMLIA, CORAL, S/KADEMLIA^[3]. Distributed hash tables provide advancement to protect against malicious attacks and used to coordinate between all the nodes interconnected in the system.



IV. HYPERMEDIA AND CAS

IPFS is a hypermedia distributed protocol. Hypermedia is the extended version of hypertext. It is a non-linear medium which contains both graphical and non-graphical information including hyperlinks. World Wide Web (WWW) is the best example of hypermedia. Hypermedia can be delivered by music players, software etc. Hypermedia allows users to click on information like images, videos etc. to link to information. It is widely termed as interactive media because of its features and functions. Hypermedia is the common term used when creating Application Programming Interfaces (APIs). The rise in cloud-based services accelerates the hypermedia APIs. It provides the power to control graphical-based services rather than traditional text-based linear services.

CAS stands for content-addressable storage. It is the method of storing data and accessing the information using the content. Traditional approach uses the concept of the storage location for storing and retrieving the information which reduces the speed and accuracy. By use of CAS, we can rapidly access the stored data with maximum accuracy. Fixed content storage and content-based storage are two different terms of same technology which mainly focuses on storage of data which do not change with time. The main feature of this is that the location of data and other related information like a number of copies etc. are unknown to the users and outside world. CAS mainly focuses on the permanent data. It is acceptable due to its quickly retrieving techniques with the data retrieved will be the same data stored.



V. APPLICATIONS

1. A mounted file system which is globally accepted.
2. Hash-key encrypted file and data sharing system.
3. A boot and root file system of a virtual machine.
4. Linked and encrypted data communication platforms.
5. Sharing and accessing a large volume of files.
6. Permanent web with no failure point.

VI. ADVANTAGES

1. It can be used for the large volume of datasets such as petabytes and much more which allows various organizations to perform computations on a large dataset for performing specified datasets.
2. It makes the process of linking and communication between a large dataset very easy and at very low cost.
3. IPFS is distributed and decentralized which make it secure and there is no point of failure of the system using the minimum number of resources.
4. Using traditional approach there is a high risk of accidental disappearance of data which is not possible in IPFS.
5. It prevents data from any malicious attacks like DDoS attacks. As the system is not centralized to attack the system one must need hundreds of supercomputers for such attacks which is not possible.

VII. DISADVANTAGES

1. IPFS uses the cryptography technique and that needs mining which requires large energy consumption.
2. Every blockchain must be digitally signed using a public-private cryptography scheme.
3. There is lack of scalability. None of the different platforms have anywhere near the throughput capacity to become a financial platform for the whole world.^[7]

VIII. CONCLUSION

This paper presents a review on IPFS and working of IPFS. Many terms which are related to IPFS are discussed. Various applications of IPFS with working principle of IPFS and future scope are mentioned. IPFS is a very challenging field which is increasing at the highest rate possible yet it is in a critical phase of its development. Future of the internet or next-generation internet depends upon the advancement of the blockchain which is the base of IPFS. Highly advance algorithms and protocols will be used to make IPFS a stable network. All the advantages and disadvantages of IPFS are presented. Despite some disadvantages, IPFS will be the future and advancement of technology can overcome the disadvantages over time. IPFS reduces the human error, cost and time by removing the centralized system. IPFS provides permanent non-failure web which is more secure and cheap then the present technology. It is the future of the file system which will occupy the traditional approach of the file system and web.

IX. REFERENCES

- [1] IPFS:en.wikipedia.org/wiki/InterPlanetary_File_System.
- [2] Website: ipfs.io.
- [3] Git: github.com/ipfs/ipfs#ipfs-papers
- [4] Git: github.com/ipfs/py-ipfs-api
- [5] Website:cloudfundoo.wordpress.com/2012/05/28/distributed-hash-tables-and-consistent-hashing.
- [6] Website:lap.umd.edu/SOC/lock42/Ch5/Ch5.html
- [7] Quora:quora.com/limitation-of-blockchain.