Paper ID – COMP – 26 – 87

Content Addressed P2P File System for the Web with Blockchain based data integrity

By –

Chaitanya Rahalkar and Dhaval Gujar

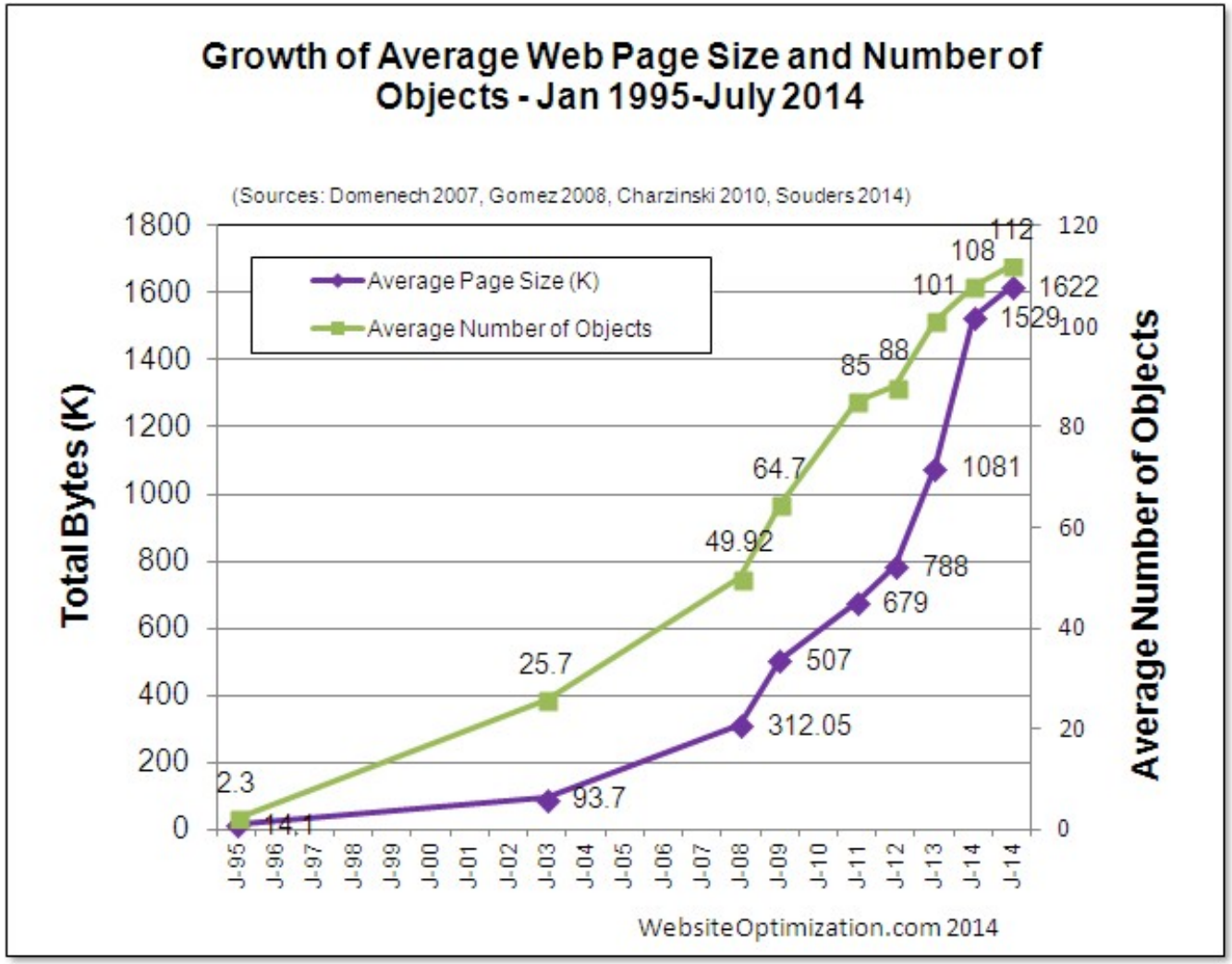
Motivation

- IPFS (InterPlanetary File System) is a proposed protocol that enhances HTTP. We are entering the era of data distribution with new challenges like:

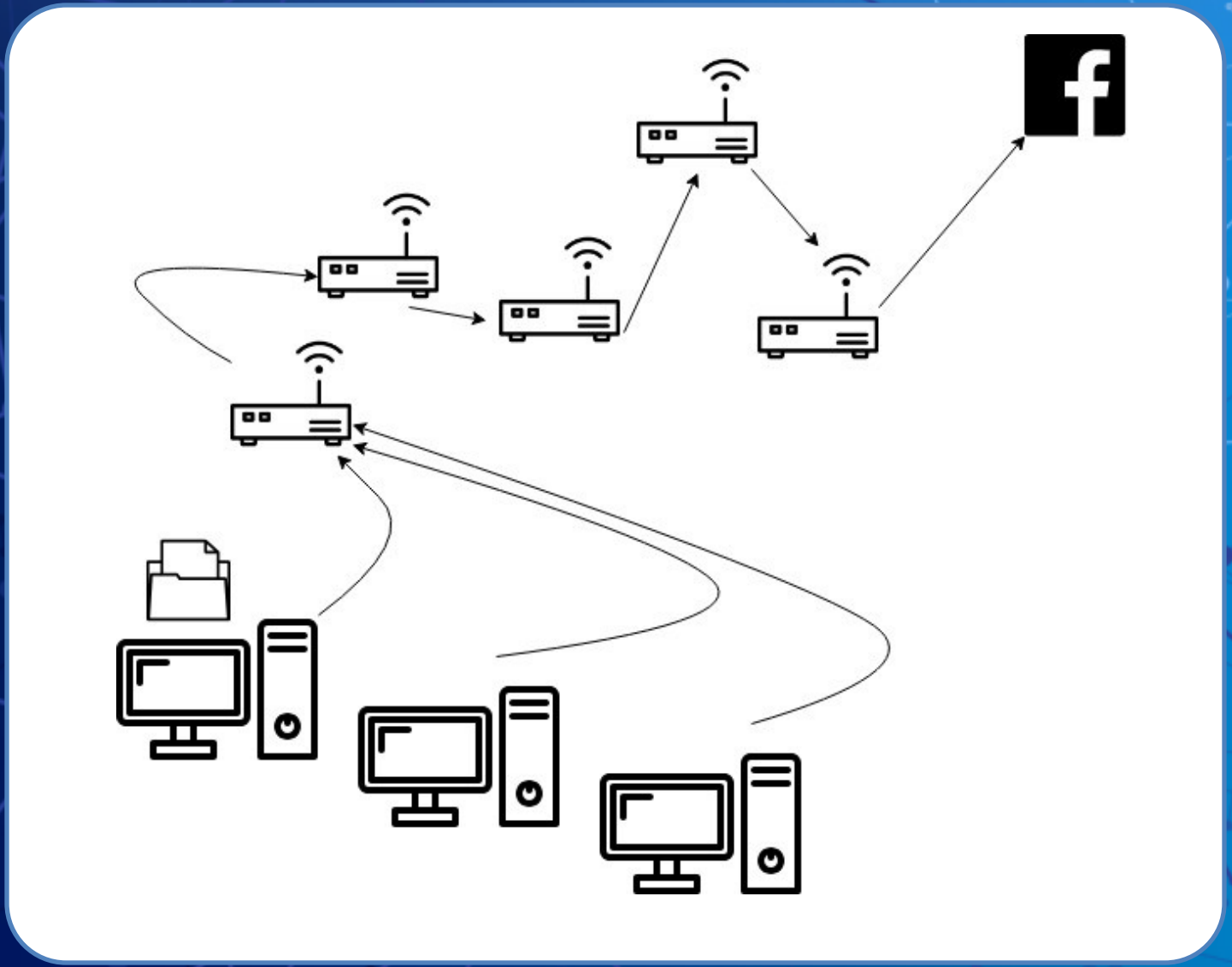
(a) Bandwidth
(b) Latency
(c) Centralisation

- To tackle all these problems, HTTP does not provide a scalable solution.
- Adding the middleware of Blockchain technology for preserving the file metadata helps to maintain the integrity of the files that are stored. Blockchain technology induces its peculiar characteristics of data integrity, data security and transparency to this file system.
- Blockchain technology is a distributed ledger system, that will preserve all the file metadata including file size, author information, checksums, date of creation and modification etc.

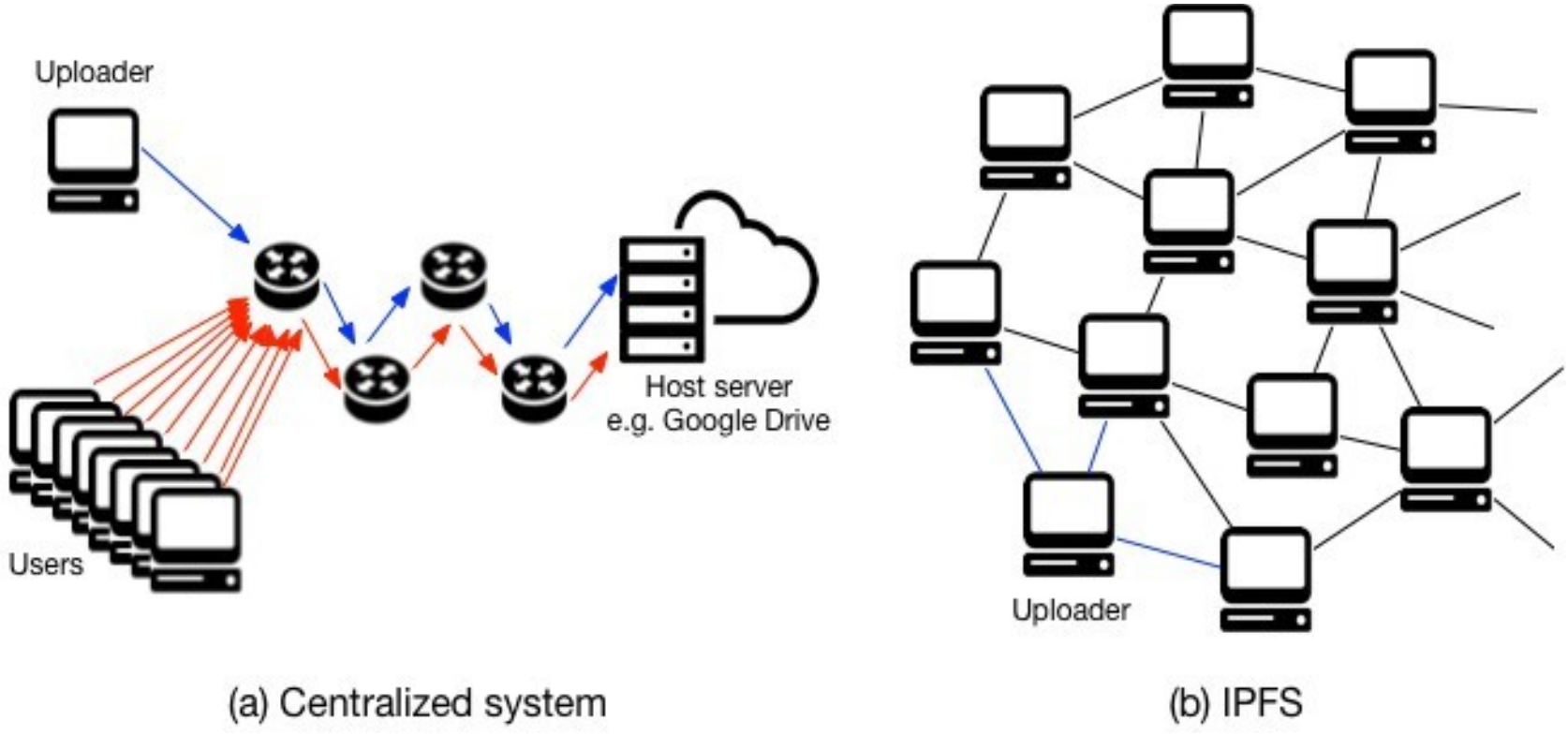
Statistics


Growth of Average Web Page Size and Number of Objects - Jan 1995-July 2014
(Sources: Domenech 2007, Gomez 2008, Charzinski 2010, Souders 2014)
Total Bytes (K)
Average Number of Objects
Average Page Size (K)
Average Number of Objects
1800
1600
1400
1200
1000
800
600
400
200
0
120
100
80
60
40
20
0
2.3
14.1
25.7
93.7
49.92
312.05
64.7
507
85
679
88
788
101
1081
108
1529
112
1622
J-95
J-96
J-97
J-98
J-99
J-00
J-01
J-02
J-03
J-04
J-05
J-06
J-07
J-08
J-09
J-10
J-11
J-12
J-13
J-14
J-14
WebsiteOptimization.com 2014

The Problem With HTTP



How IPFS Fixes It



(a) Centralized system

(b) IPFS

Terminologies of The Proposed Model

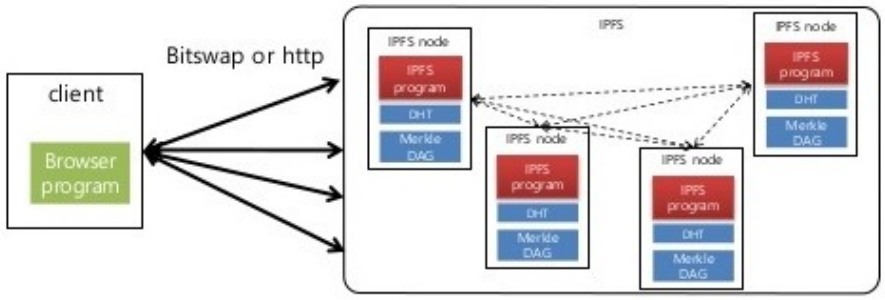
- Content Addressed File System
- Distributed Hash Tables
- Blockchain Technology
- P2P Decentralised System

What is Content Addressed File System?

- Everything on the World Wide Web is addressed with a URL that maps to the location of the file on the Internet. The IP address assigned to a website locates the file on the WWW.

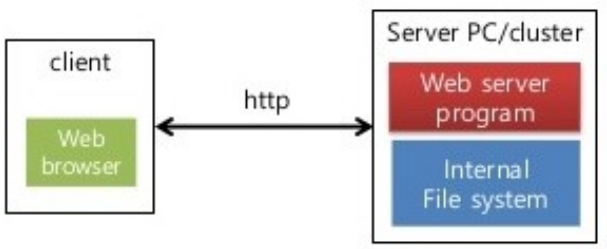
- However, in a *Content Addressed File System*, the file is accessed just on the basis of its content and not on its location.

ipfs/QmWATWQ7fVPP2EFGu71UkfnqhYXDYH566qy47CnJDgvs8u



Content Addressed

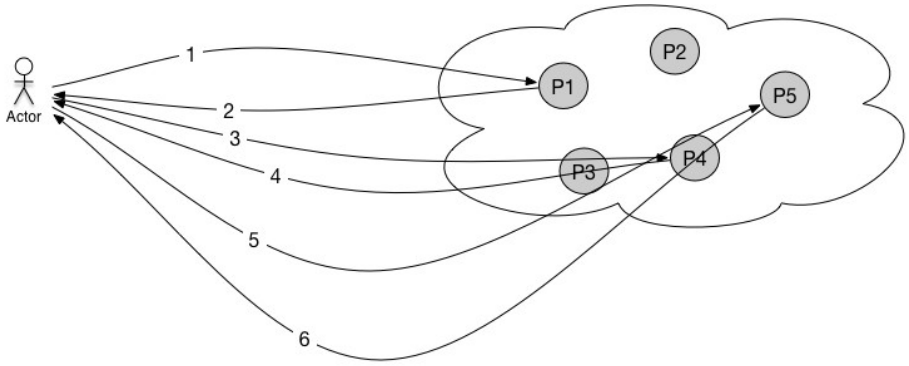
http://10.20.30.40/foo/bar/baz.png



Location Addressed

Distributed Hash Tables in IPFS Model

- A DHT is simply a key-value store distributed across a number of nodes in a network. The keys are distributed among nodes with a deterministic algorithm. Each node is responsible for a portion of the hash table.
- A routing algorithm allows to perform requests in the hash table without knowing every node of the network.
- A DHT is very scalable because the data are uniformly distributed among nodes and lookup time generally grows in O(log(N)).
- Without storing the entire routing table of the network (the addresses of each nodes). Basically you ask the closest node to the data identifier you know which itself asks the closest node it knows and so on reducing the size of the jump at each step.



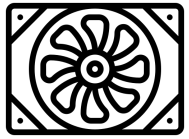
1. STORE “MyKey” / “My Value”
2. I’m not responsible for “MyKey” - but P4 is closer
3. STORE “MyKey” / “My Value”
4. I’m not responsible for “MyKey” - but P5 is closer
5.STORE “MyKey” / “My Value”
6. OK - value is stored.

1. GET “MyKey”
2. I’m not responsible for “MyKey” - but P4 is closer
3. GET “MyKey”
4. I’m not responsible for “MyKey” - but P5 is closer
5.GET “MyKey”
6. OK - here is “My Value”

Need of P2P Technology

- Hosting Petabytes of data.

- Computing large data over distributed organisations.



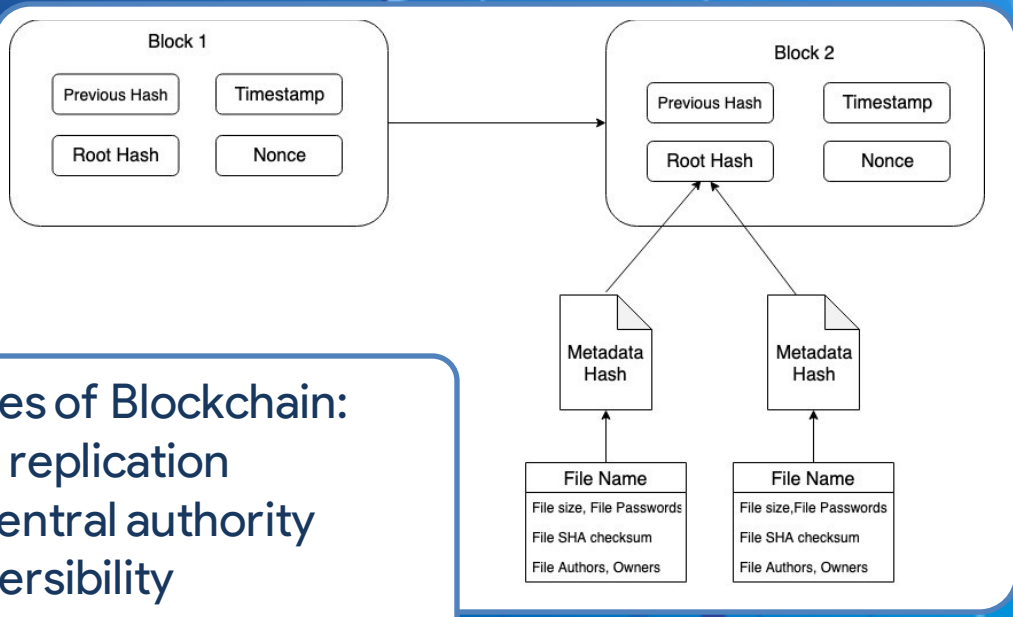
- High volume, high definition on demand, real time streaming of data.

- Versioning and linking off massive datasets.

- Preventing accidental disappearances of important files.

The Purpose of Blockchain

- Blockchain technology is a distributed ledger system, that will preserve all the file metadata including file size, author information, checksums, date of creation and modification etc.
- In a Blockchain, each node of the network stores the full data. So it is not the same idea as the DHT in which data are divided among nodes.



Properties of Blockchain:

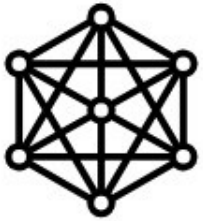
- Data replication
- No central authority
- Irreversibility
- Accessibility
- Time stamping

Id	File Creation Date	IPFS Hash Value	File Access Date	File Access Time	IPFS hash of modified file	Other metadata
1	12/06/18	Qm....2	14/06/18	12:45 PM	Qm....9	--
2	16/06/18	Qm....19	20/06/18	1:05 PM	Qm....25	--

Table 1: Example Record

How does the model work?

Each file and all of the blocks within it are given a unique fingerprint called a cryptographic hash.



The file is uploaded on the IPFS distributed network.

DISTRIBUTED LEDGER

File metadata is sent to the Ethereum Blockchain.

The file integrity is checked by retrieving the metadata of the file from the Blockchain. The file is identified by its hash and is retrieved from the network.

Application Scope

- With the current demands of the industry, IPFS and Blockchain is a perfect pair to perform scalable and fault tolerant tasks.

Following are some of the tasks that can be achieved:

- **Preservation of Massive Datasets:** With the distributed technology, the model allows people to store large datasets, showing fast performance with decentralised archiving system. Along with that, the integrity of the datasets can be preserved.
- **Sensitive Data Storage:** Sensitive government documents, user data like Aadhaar cards, bond papers, contracts etc. can be securely and safely stored with this model avoiding fraud cases. Repudiation is completely avoided.
- **Content Delivery:** Secured P2P content delivery saves millions in bandwidth, also providing better performance.

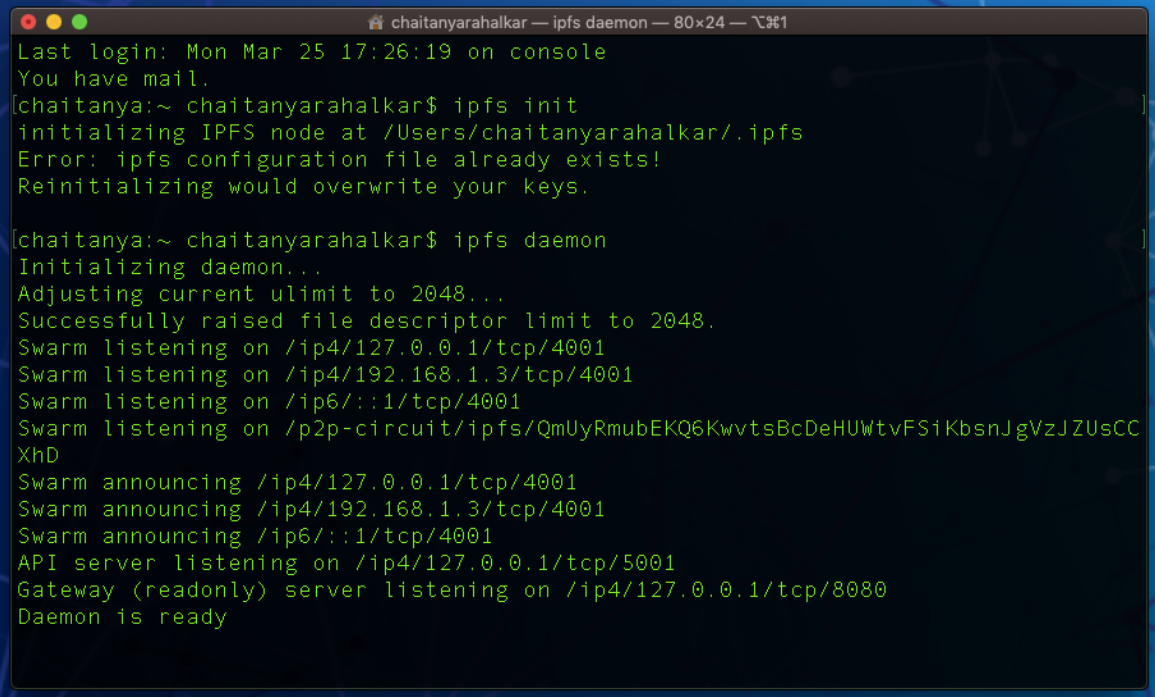
Conclusion

- A secured and integrity compliant system was proposed using the P2P feature of IPFS and the tamperproof principle of Blockchain technology.
- This model is a complete solution to the various problems faced by HTTP and data security.
- With minimal hardware requirements, any node in the decentralised network can serve data, improving bandwidth, latency and availability.
- The four main components namely DHTs, Blockchain, P2P Networks and Content Addressed File System together make the model a secured, reliable and fault tolerant system.

“Without engineers, science is just philosophy.”

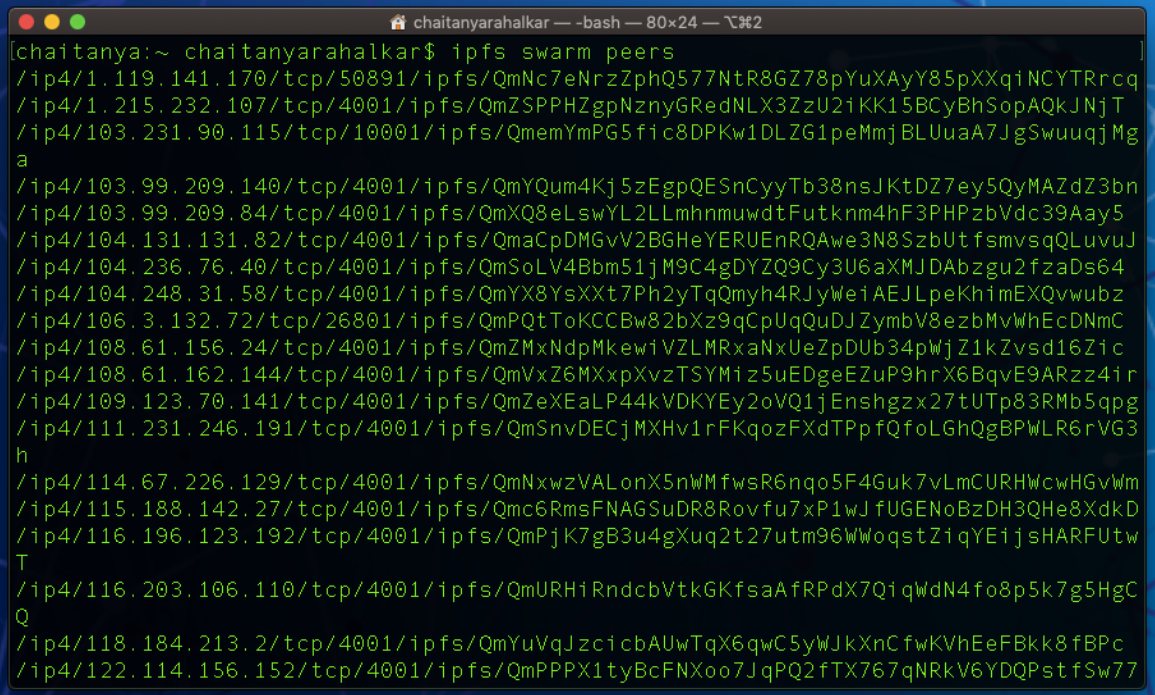
THANK YOU!

Implementation



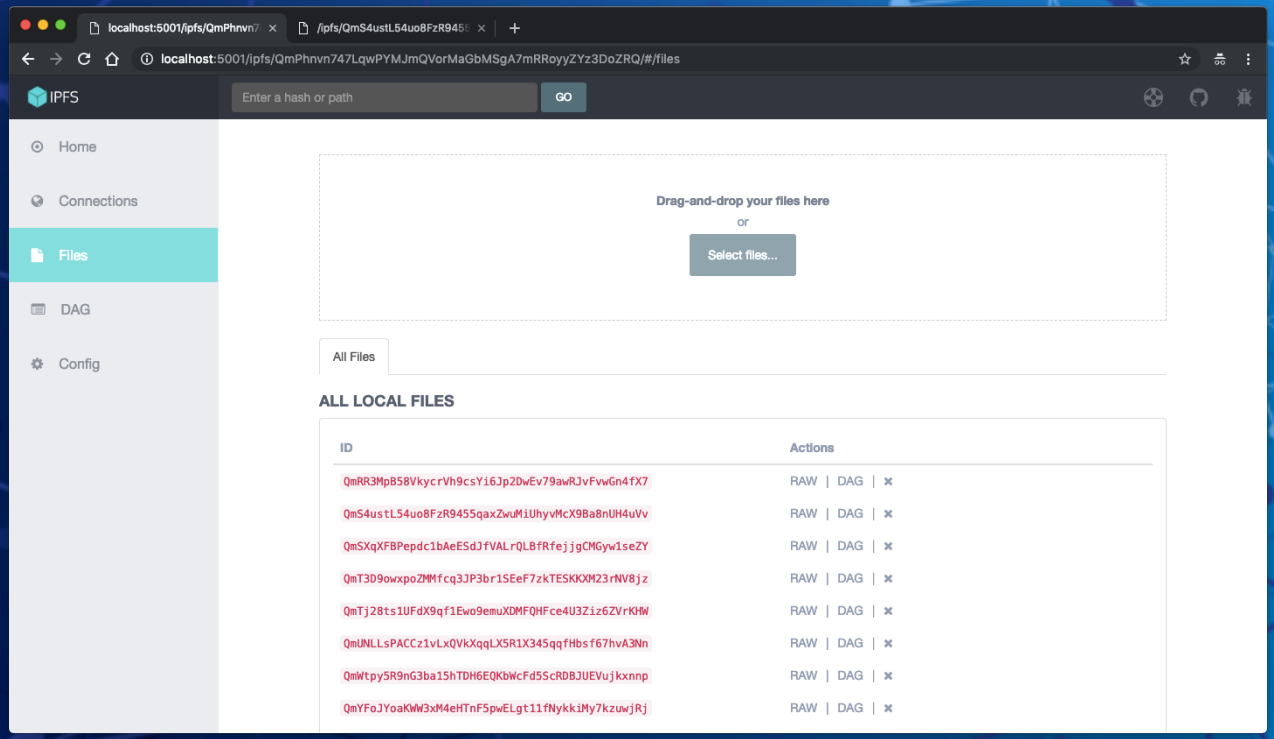
In order to become a part of the P2P network, we create an IPFS node. The IPFS node creates a TCP socket and connects to all the peers in the network. Instead of an IP address, the node is identified by a unique hash value that is generated from the public-private key pair.

Implementation



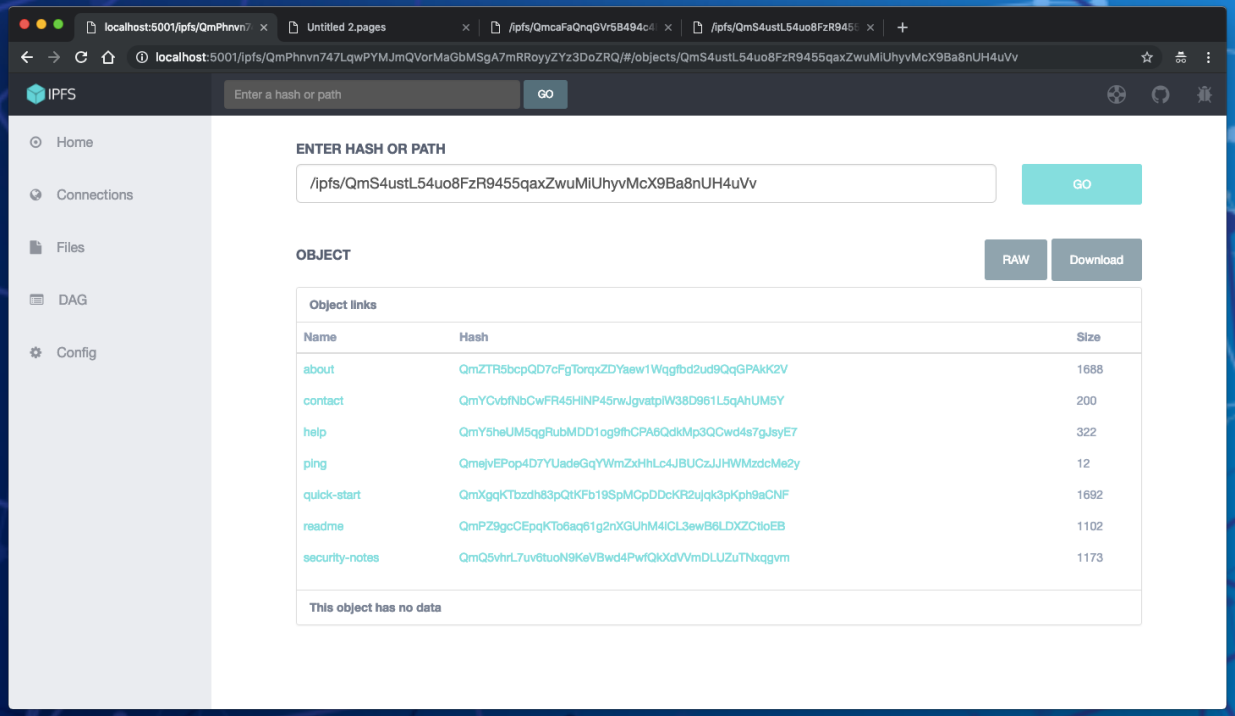
After creating an IPFS node, we connect to all the peers in the P2P network.

Implementation



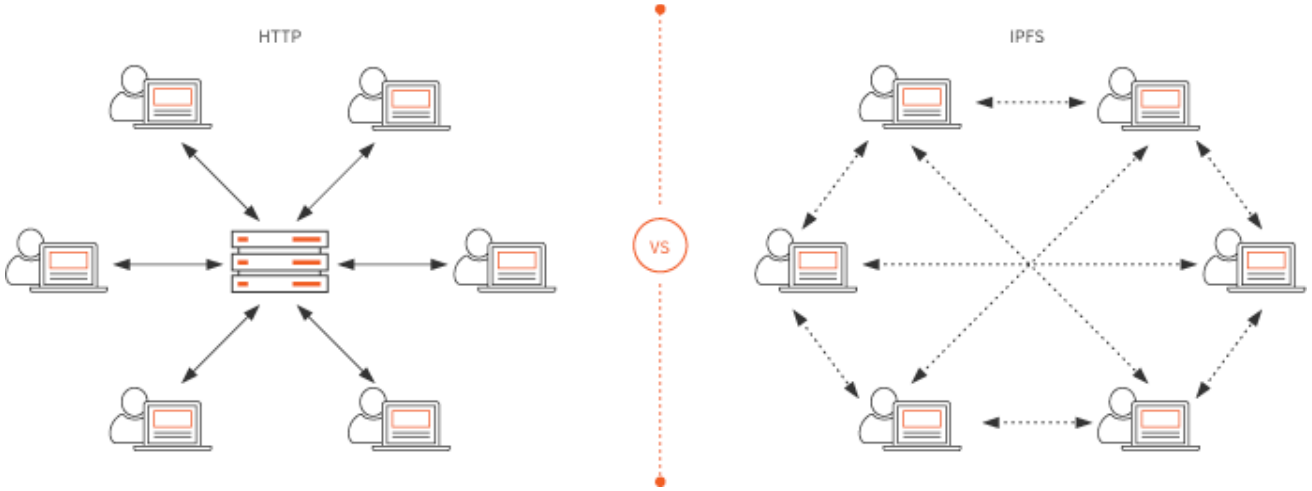
After connecting to the IPFS network, the web interface allows us to upload files to the P2P network. A unique hash prefixed with *'Qm'* indicates that it is an IPFS hash.

Implementation



Entering the unique IPFS hash corresponding to a file, allows us to retrieve the file from the P2P network.

P2P Technology vs Client Server Model



- The HTTP protocol is based on the Client-Server model, while IPFS is a P2P protocol involving no centralised system.
- With no control of any central authority, data integrity and safety is maintained in a decentralised system.

References i

J. Benet.
IPFS - Content Addressed, Versioned, P2P File System.
[Cs] arXiv.org, 2014.

J. Benet.
Ipfs - content addressed, versioned, p2p file system.
ArXiv, abs/, 1407:3561, 2014.

M. Kelly et al.
Interplanetary wayback: Peer-to-peer permanence of web archives.
In N. F. al. and, editor, *Research and Advanced Technology for Digital Libraries*, pages 411–16. Springer, International Publishing, 2016.

References ii


P. Labs.
IPFS Is the Distributed Web.
IPFS.


B. Marr.
How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read.
Forbes.


S. Nakamoto.
A Peer-to-Peer Electronic Cash System, Bitcoin, 2009.


H. Wang, Z. Zheng, S. Xie, H.-N. Dai, and X. Chen.
Blockchain challenges and opportunities: a survey,
10(1504):352–375, 2018.


T. E. Wiki.
GitHub, 2019.

References iii

G. Zyskind et al.

Decentralizing privacy: Using blockchain to protect personal data.

2015:180–84, 2015.