A Seminar On

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

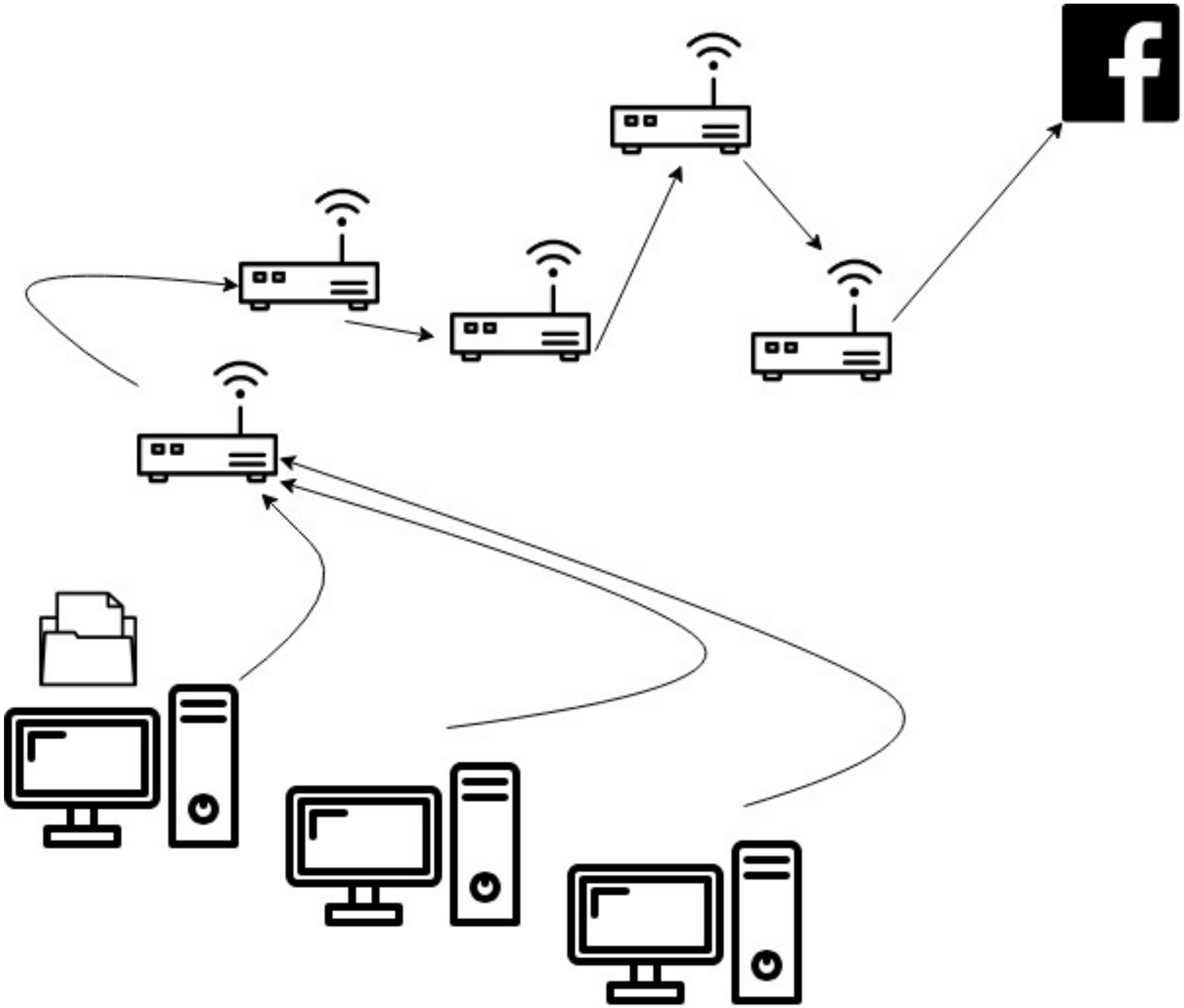
Domain- Information Security
Presented By -
Chaitanya Rahalkar

Motivation

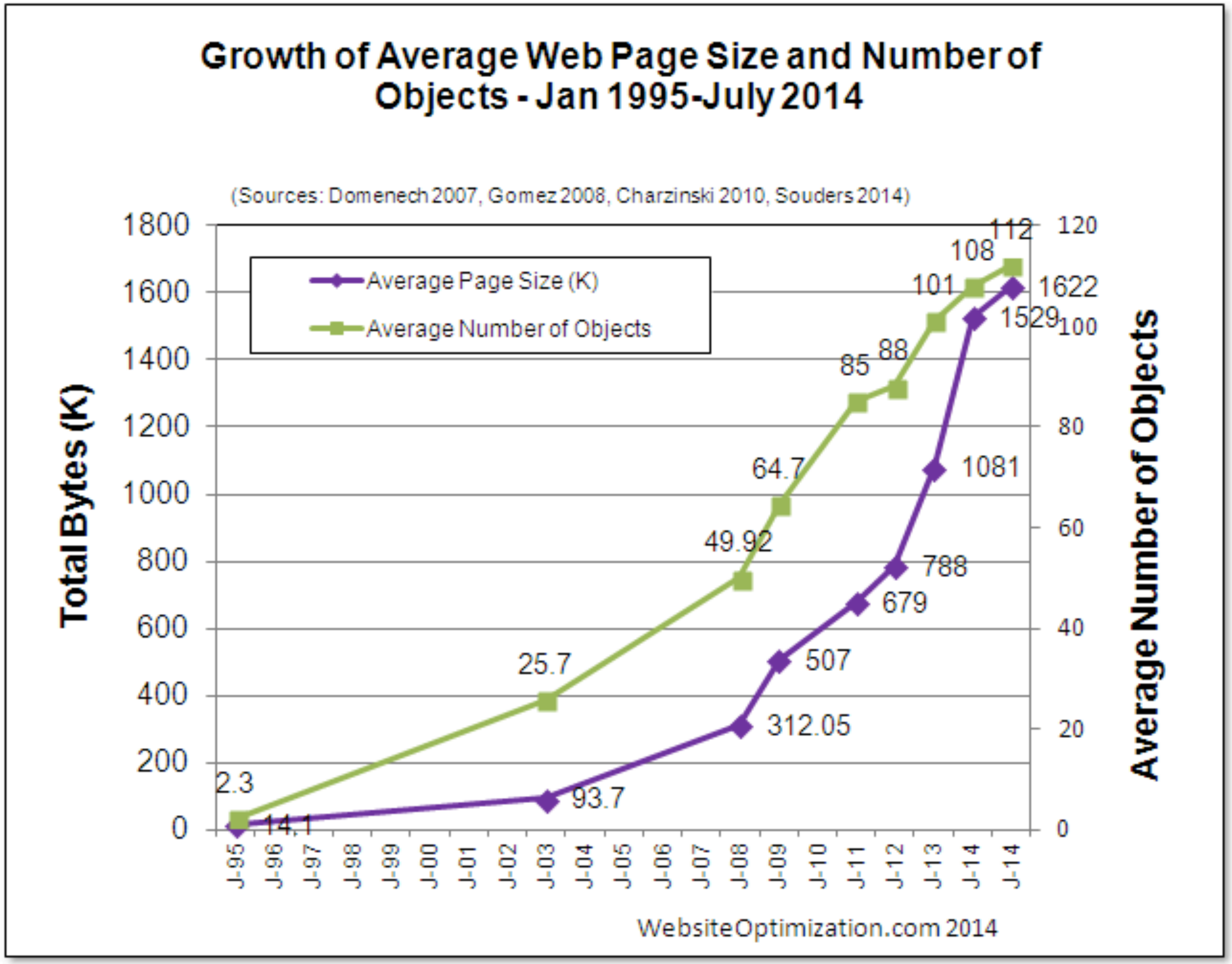
IPFS(InterPlanetary File System) is a proposed protocol that enhances HTTP. We are entering the era of data distribution with new challenges like:
(a) Bandwidth
(b) Latency
(c) Centralisation
To tackle all these problems, HTTP does not provide a scalable solution.

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.

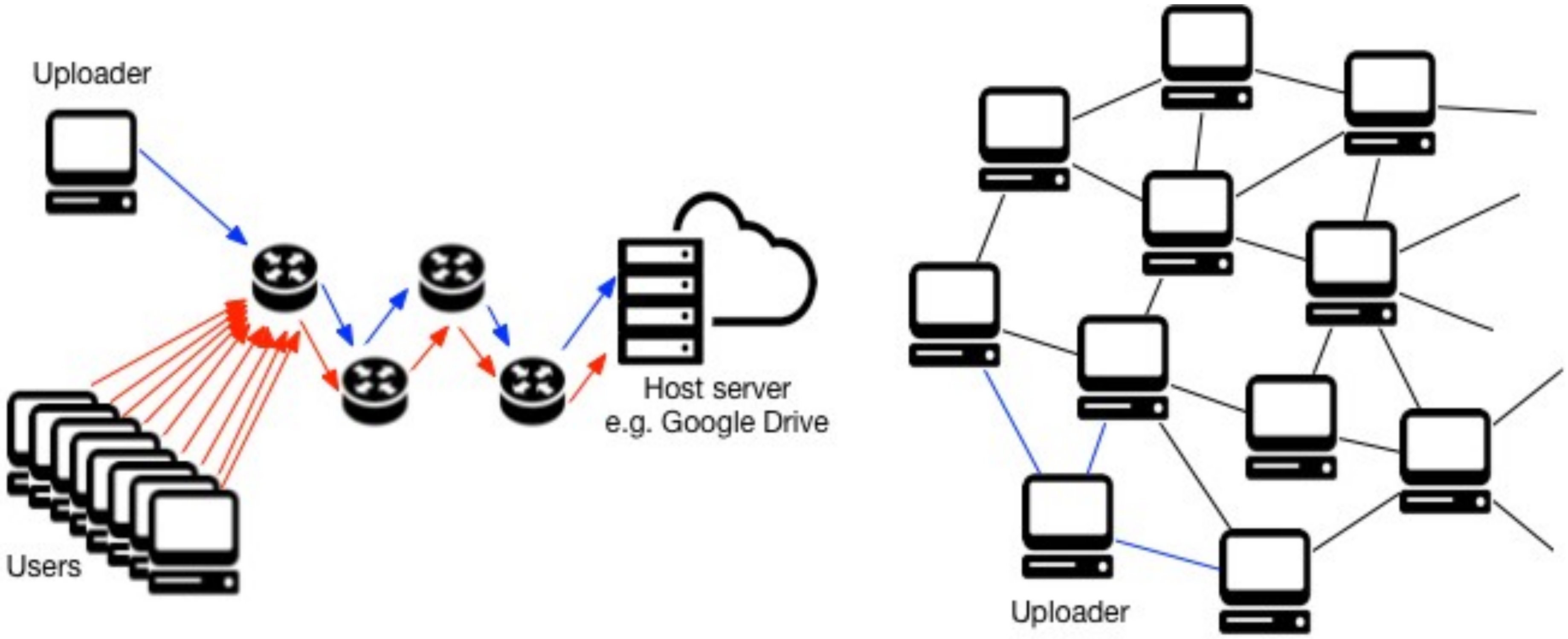
The Problem With HTTP



Statistics


Growth of Average Web Page Size and Number of Objects - Jan 1995-July 2014
(Sources: Domenech 2007, Gomez 2008, Charzinski 2010, Souders 2014)
Average Page Size (K)
Average Number of Objects
Total Bytes (K)
Average Number of Objects
1800
1600
1400
1200
1000
800
600
400
200
0
120
100
80
60
40
20
0
2.3
25.7
49.92
64.7
85
88
101
108
112
14.1
93.7
312.05
507
679
788
1081
1529
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

How IPFS Fixes It



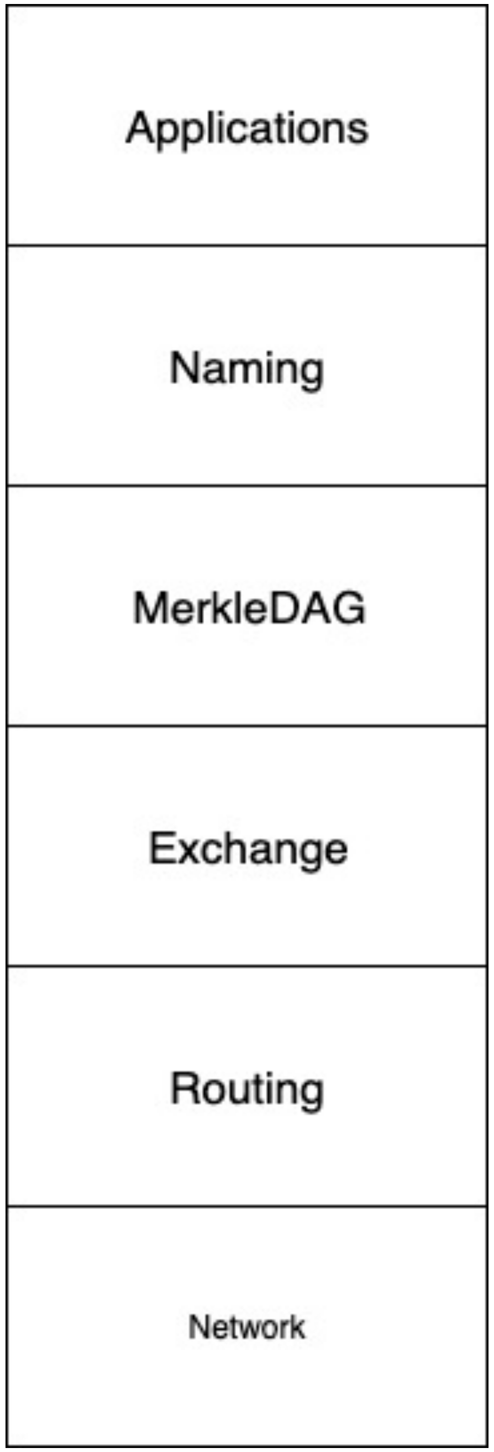
(a) Centralized system

(b) IPFS

Terminologies Of The Proposed Model

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

IPFS Architecture Stack



Applications

Naming

MerkleDAG

Exchange

Routing

Network

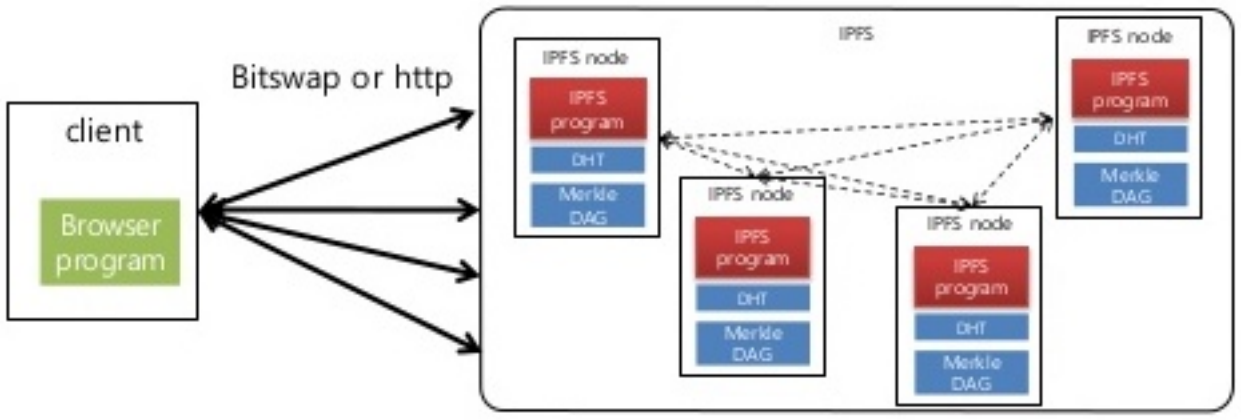
What is Content Addressed File System?

- 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.

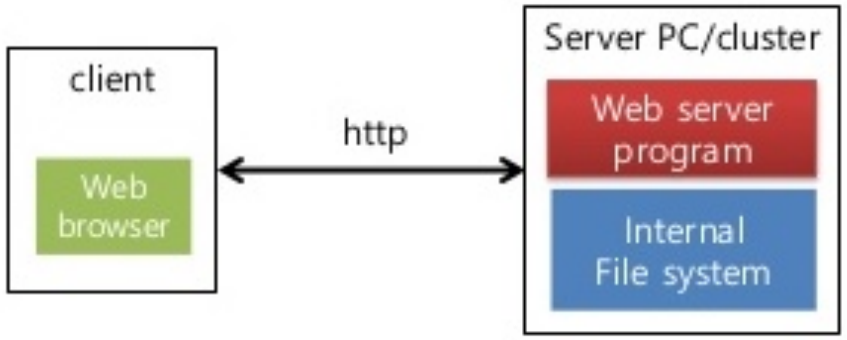
- However, in a *Content Addressed File System,* the file is accessed just on the basis of its content and not on its location.

ipfs/QmWATWQ7fVPP2EFGu71UkfnqhYXDYH566qy47CnJDgvs8u

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



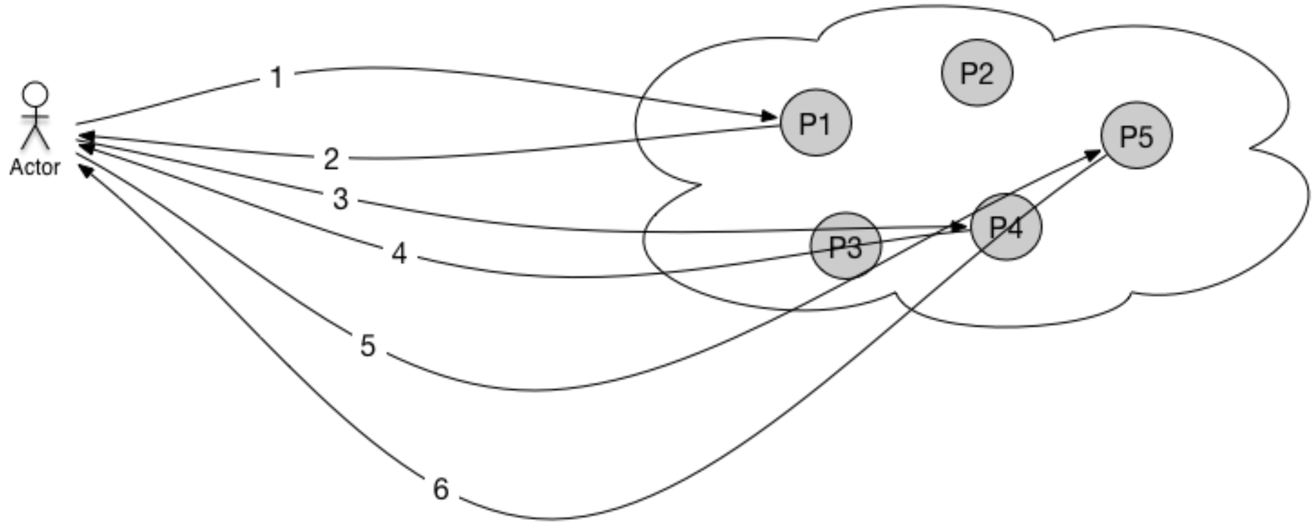
Content Addressed



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.

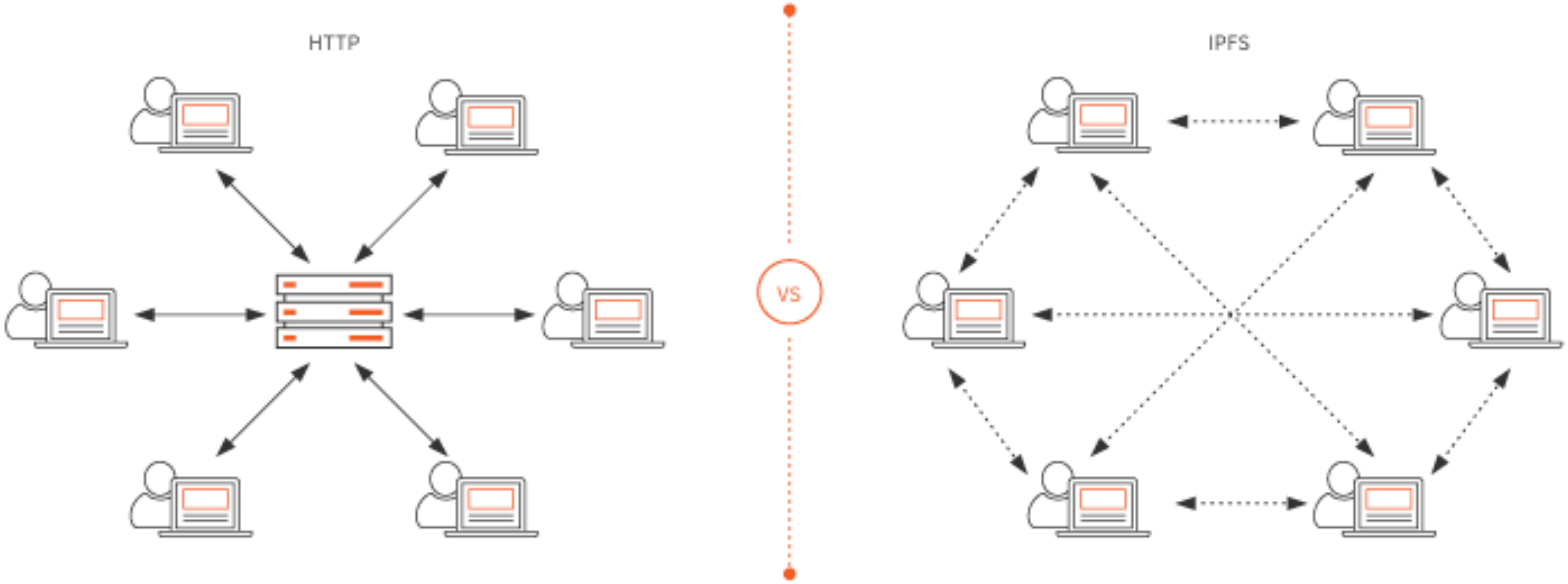
1. GET “MyKey”
2. I’m not responsible for “MyKey” - but P4 is closer
3. GET “MyKey”
4. I’m not responsible for “MyKey” - but P5 is closer
5.GET “MyKey”
6. OK - here is “My Value”

Iterative Lookup in DHT

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.

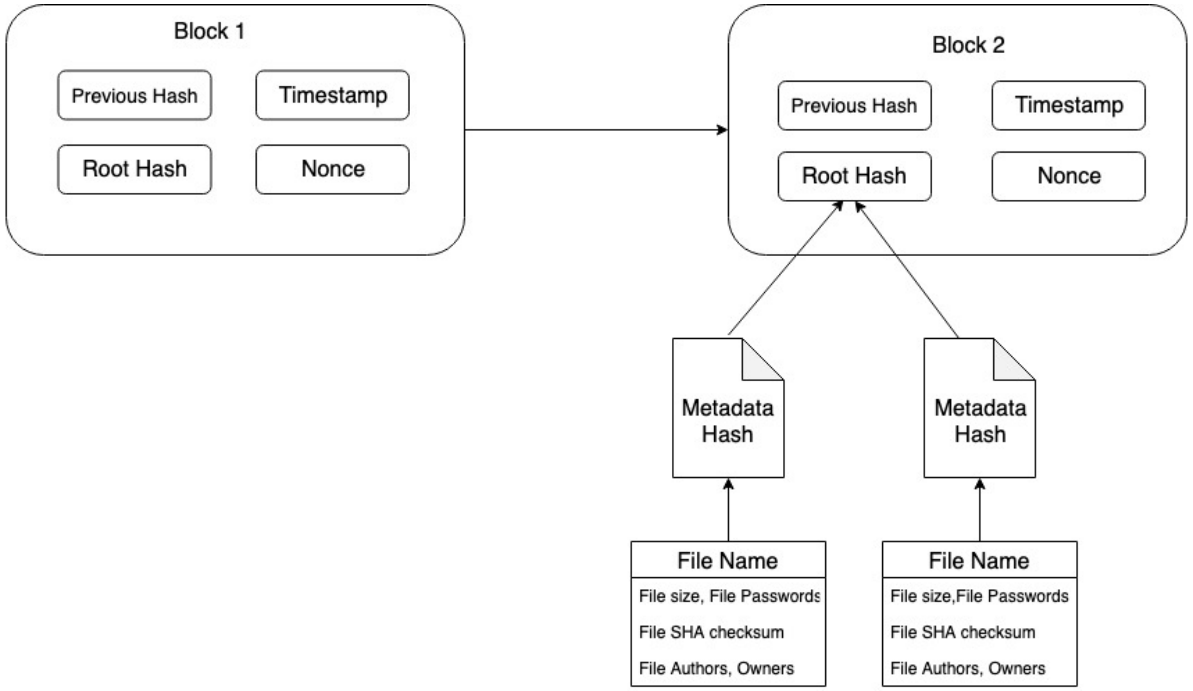
P2P Technology vs Client Server Model



- 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.

The Purpose Of Blockchain

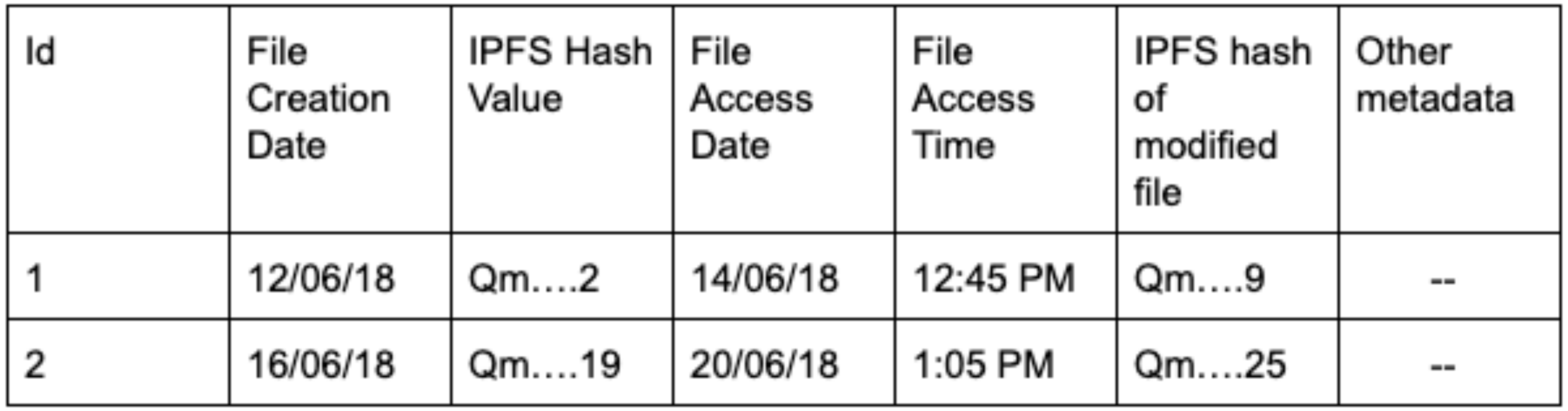


- Blockchain technology is a distributed ledger system, that will preserve all the file metadata including file size, author information, checksums, date of creation and modification etc.

- In a blockchain, each node of the network stores the full data. So it is absolutely not the same idea as the DHT in which data are divided among nodes.

- Properties Of Blockchain:

1. Data replication
2. No central authority
3. Irreversibility
4. Accessibility
5. 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 the model works?

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.

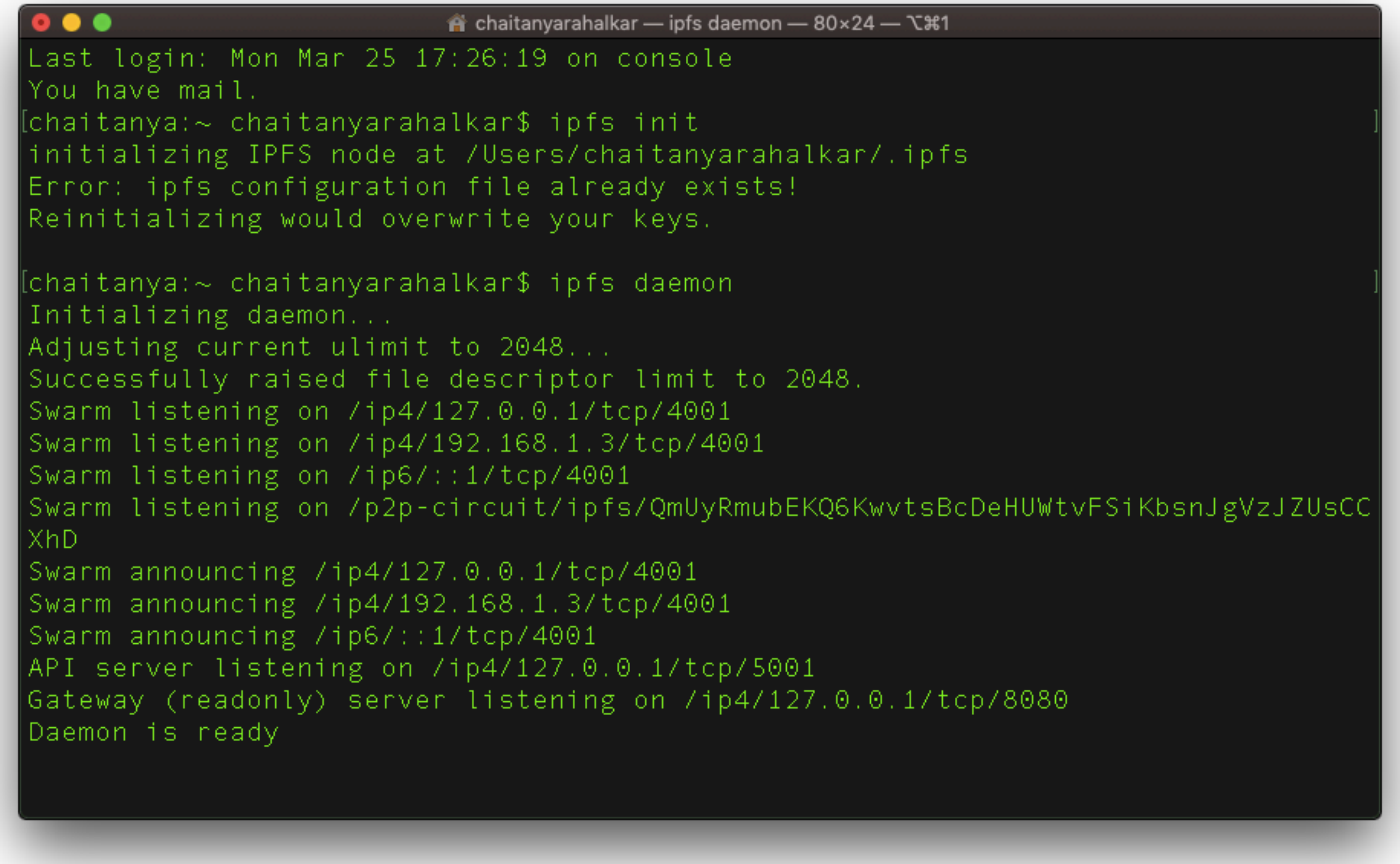
Following are some of the tasks that can be achieved:

1. 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.

2. Sensitive Data Storage: Sensitive government documents, user data like Adhaar cards, bond papers, contracts etc. can be securely and safely stored with this model avoiding fraud cases. Repudiation is completely avoided.

3. Content Delivery: Secured P2P content delivery saves millions in bandwidth, also providing better performance.

Implementation

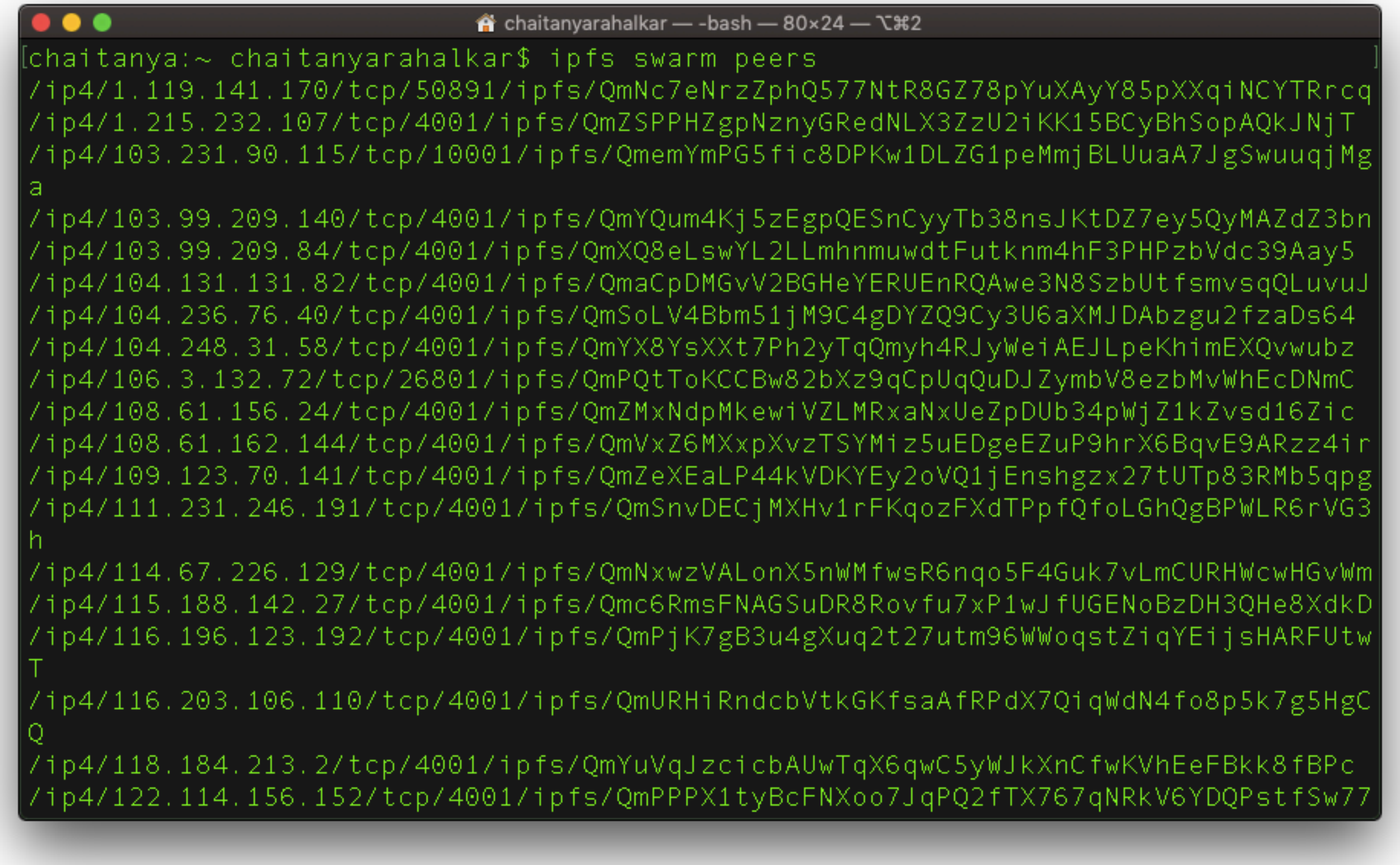


```
Last login: Mon Mar 25 17:26:19 on console
You have mail.
[chaitanya:~ chaitanyarahalkar$ ipfs init                                        ]
initializing IPFS node at /Users/chaitanyarahalkar/.ipfs
Error: ipfs configuration file already exists!
Reinitializing would overwrite your keys.

[chaitanya:~ chaitanyarahalkar$ ipfs daemon                                      ]
Initializing daemon...
Adjusting current ulimit to 2048...
Successfully raised file descriptor limit to 2048.
Swarm listening on /ip4/127.0.0.1/tcp/4001
Swarm listening on /ip4/192.168.1.3/tcp/4001
Swarm listening on /ip6/::1/tcp/4001
Swarm listening on /p2p-circuit/ipfs/QmUyRmubEKQ6KwvtsBcDeHUWtvFSiKbsnJgVzJZUsCC
XhD
Swarm announcing /ip4/127.0.0.1/tcp/4001
Swarm announcing /ip4/192.168.1.3/tcp/4001
Swarm announcing /ip6/::1/tcp/4001
API server listening on /ip4/127.0.0.1/tcp/5001
Gateway (readonly) server listening on /ip4/127.0.0.1/tcp/8080
Daemon is ready
```

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.

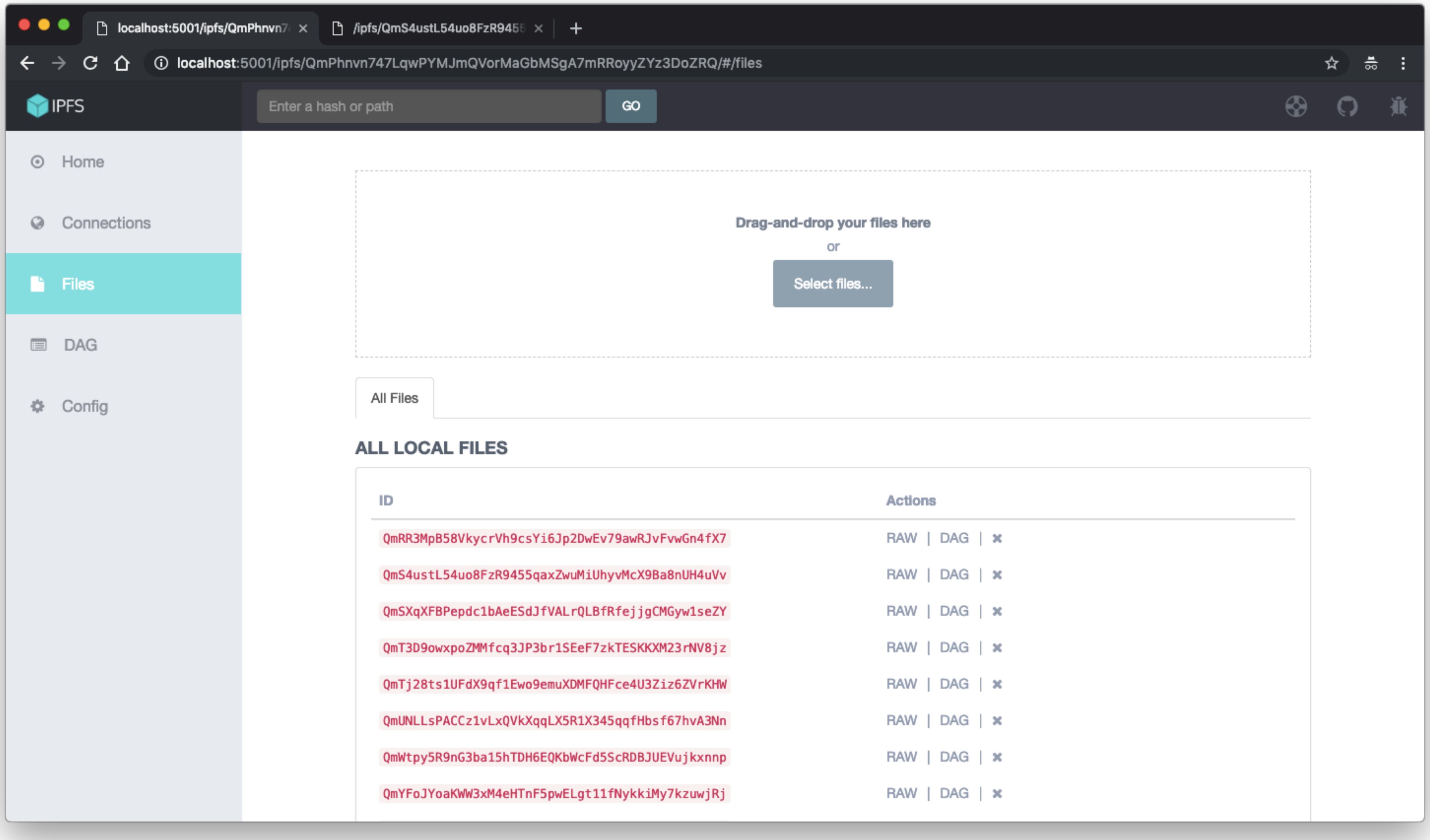
Implementation



```
[chaitanya:~ chaitanyarahalkar$ ipfs swarm peers
/ip4/1.119.141.170/tcp/50891/ipfs/QmNc7eNrzZphQ577NtR8GZ78pYuXAyY85pXXqiNCYTRrcq
/ip4/1.215.232.107/tcp/4001/ipfs/QmZSPPHZgpNznyGRedNLX3ZzU2iKK15BCyBhSopAQkJNjT
/ip4/103.231.90.115/tcp/10001/ipfs/QmemYmPG5fic8DPKw1DLZG1peMmjBLUuaA7JgSwuuqjMg
a
/ip4/103.99.209.140/tcp/4001/ipfs/QmYQum4Kj5zEgpQESnCyyTb38nsJKtDZ7ey5QyMAZdZ3bn
/ip4/103.99.209.84/tcp/4001/ipfs/QmXQ8eLswYL2LLmhnmuwdtFutknm4hF3PHPzbVdc39Aay5
/ip4/104.131.131.82/tcp/4001/ipfs/QmaCpDMGvV2BGHeYERUEnRQAwe3N8SzbUtfsmvsqQLuvuJ
/ip4/104.236.76.40/tcp/4001/ipfs/QmSoLV4Bbm51jM9C4gDYZQ9Cy3U6aXMJDAbzgu2fzaDs64
/ip4/104.248.31.58/tcp/4001/ipfs/QmYX8YsXXt7Ph2yTqQmyh4RJyWeiAEJLpeKhimEXQvwubz
/ip4/106.3.132.72/tcp/26801/ipfs/QmPQtToKCCBw82bXz9qCpUqQuDJZymbV8ezbMvWhEcDNmC
/ip4/108.61.156.24/tcp/4001/ipfs/QmZMxNdpMkewiVZLMRxaNxUeZpDUb34pWjZ1kZvsd16Zic
/ip4/108.61.162.144/tcp/4001/ipfs/QmVxZ6MXxpXvzTSYMiz5uEDgeEZuP9hrX6BqvE9ARzz4ir
/ip4/109.123.70.141/tcp/4001/ipfs/QmZeXEaLP44kVDKYEy2oVQ1jEnshgzx27tUTp83RMb5qpg
/ip4/111.231.246.191/tcp/4001/ipfs/QmSnvDECjMXHv1rFKqozFXdTPpfQfoLGhQgBPWLR6rVG3
h
/ip4/114.67.226.129/tcp/4001/ipfs/QmNxwzVALonX5nWMfwsR6nqo5F4Guk7vLmCURHWcwHGvWm
/ip4/115.188.142.27/tcp/4001/ipfs/Qmc6RmsFNAGSuDR8Rovfu7xP1wJfUGENoBzDH3QHe8XdkD
/ip4/116.196.123.192/tcp/4001/ipfs/QmPjK7gB3u4gXuq2t27utm96WWoqstZiqYEijsHARFUtw
T
/ip4/116.203.106.110/tcp/4001/ipfs/QmURHiRndcbVtkGKfsaAfRPdX7QiqWdN4fo8p5k7g5HgC
Q
/ip4/118.184.213.2/tcp/4001/ipfs/QmYuVqJzcicbAUwTqX6qwC5yWJkXnCfwKVhEeFBkk8fBPc
/ip4/122.114.156.152/tcp/4001/ipfs/QmPPPX1tyBcFNXoo7JqPQ2fTX767qNRkV6YDQPstfSw77
```

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

Implementation



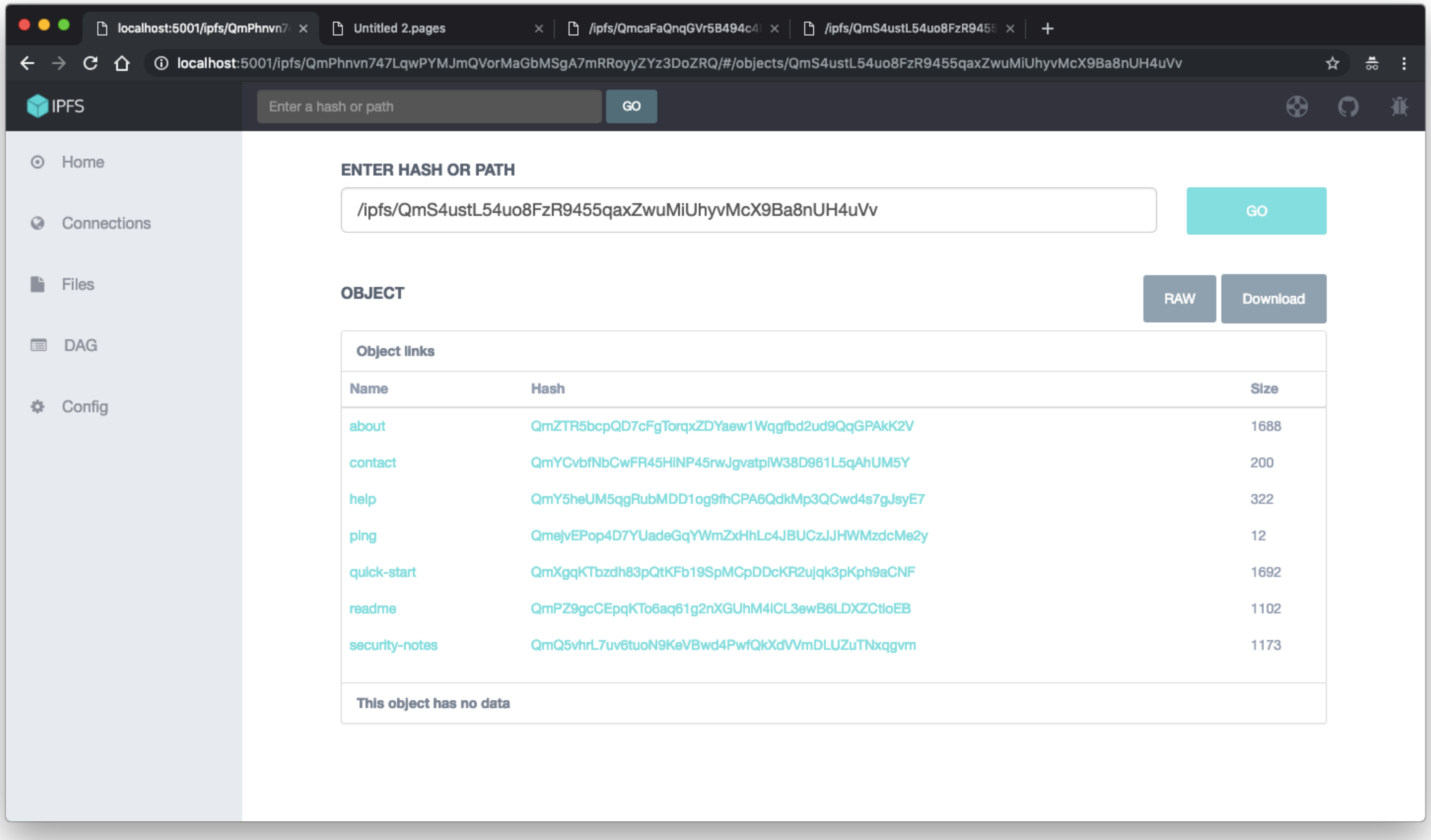
localhost:5001/ipfs/QmPhnvn7
/ipfs/QmS4ustL54uo8FzR9455
localhost:5001/ipfs/QmPhnvn747LqwPYMJmQVorMaGbMSgA7mRRoyyZYz3DoZRQ/#/files
IPFS
Enter a hash or path
GO
Home
Connections
Files
DAG
Config
Drag-and-drop your files here
or
Select files...
All Files

ALL LOCAL FILES

ID	Actions
QmRR3MpB58VkycrVh9csYi6Jp2DwEv79awRJvFvwGn4fX7	RAW | DAG | ✖
QmS4ustL54uo8FzR9455qaxZwuMiUhyvMcX9Ba8nUH4uVv	RAW | DAG | ✖
QmSXqXFBPepdc1bAeESdJfVALrQLBfRfejjgCMGyw1seZY	RAW | DAG | ✖
QmT3D9owxpoZMMfcq3JP3br1SEeF7zkTESKKXM23rNV8jz	RAW | DAG | ✖
QmTj28ts1UFdX9qf1Ewo9emuXDMFQHFce4U3Ziz6ZVrKHW	RAW | DAG | ✖
QmUNLLsPACCz1vLxQVkXqqLX5R1X345qqfHbsf67hvA3Nn	RAW | DAG | ✖
QmWtpy5R9nG3ba15hTDH6EQKbWcFd5ScRDBJUEVujkxnnp	RAW | DAG | ✖
QmYFoJYoaKWW3xM4eHTnF5pwELgt11fNykkiMy7kzuwjRj	RAW | DAG | ✖

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



localhost:5001/ipfs/QmPhnvn7
Untitled 2.pages
/ipfs/QmcaFaQnqGVr5B494c4
/ipfs/QmS4ustL54uo8FzR9455
localhost:5001/ipfs/QmPhnvn747LqwPYMJmQVorMaGbMSgA7mRRoyyZYz3DoZRQ/#/objects/QmS4ustL54uo8FzR9455qaxZwuMiUhyvMcX9Ba8nUH4uVv
IPFS
Enter a hash or path
GO
Home
Connections
Files
DAG
Config

ENTER HASH OR PATH

/ipfs/QmS4ustL54uo8FzR9455qaxZwuMiUhyvMcX9Ba8nUH4uVv

OBJECT

Object links

Name	Hash	Size
about	QmZTR5bcpQD7cFgTorqxZDYaew1Wqgfbd2ud9QqGPAkK2V	1688
contact	QmYCvbfNbCwFR45HiNP45rwJgvatpiW38D961L5qAhUM5Y	200
help	QmY5heUM5qgRubMDD1og9fhCPA6QdkMp3QCwd4s7gJsyE7	322
ping	QmejvEPop4D7YUadeGqYWmZxHhLc4JBUCzJJHWMzdcMe2y	12
quick-start	QmXgqKTbzdh83pQtKFb19SpMCpDDcKR2ujqk3pKph9aCNF	1692
readme	QmPZ9gcCEpqKTo6aq61g2nXGUhM4iCL3ewB6LDXZCtioEB	1102
security-notes	QmQ5vhrL7uv6tuoN9KeVBwd4PwfQkXdVVmDLUZuTNxqgvm	1173

This object has no data

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

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.

Future Enhancements

- Allowing Asymmetric key encryption storage schemes like PGP to store encrypted data on the IPFS network.
- Creating a full fledged web application to allow interfacing with the data.
- Adding a blockchain interactive web interface.

References

- Benet, Juan. (2014). IPFS - Content Addressed, Versioned, P2P File System.
- Nakamoto, Satoshi. (2009). Bitcoin: A Peer-to-Peer Electronic Cash System. Cryptography [2]
- https://acadpubl.eu/hub/2018-119-15/4/751.pdf
- https://medium.com/@mycoralhealth/learn-to-securely-share-files-on-the-blockchain-with-ipfs-219ee47df54c
- https://en.wikipedia.org/wiki/Distributed_hash_table
- https://medium.com/@michael.dufel_10220/distributed-hash-tables-and-why-they-are-better-than-blockchain-for-exchanging-health-records-d469534cc2a5
- https://medium.com/coinmonks/ipfs-blockchain-decentralised-file-storage-9ef3a1fa307b
- https://hackernoon.com/ipfs-a-complete-analysis-of-the-distributed-web-6465ff029b9b
- https://www.cio.com/article/3174144/the-permanent-web-for-healthcare-with-ipfs-and-blockchain.html

Thank You