A blockchain-based web hosting scheme for content-addressed distributed file system

I. Related work

The emergence of the internet has solved a lot of problems of the world like data availability, data distribution, communication, etc. But today's infrastructure of the internet does not reflect the decentralization, heterogeneity, and privacy of consumers that was originally proposed by the founder of internet Tim Berners Lee, in his 1989 www whitepaper[1]. This creates many challenges such as central point of failure, the overhead of intermediation leading to extra cost both financially and with time, fraud, privacy violation, censorship by political and other entities, etc. One of the approaches that can be taken is to switch from client-server architecture to a peer-to-peer architecture. However, this leads to new technical challenges including some new security vulnerability. A promising technology in this regard could be what was proposed in 2008 called blockchain by the name Satoshi Nakamoto proposed using timestamps of transactions by hashing them into an ongoing chain of hash-based proof-of-work. [2] This idea was implemented to a cryptocurrency in bitcoin but later integrated to other versatile areas. In the field of file storage blockchain technology promises a huge improvement over the traditional centralized file storage system because of no single point failure, disintermediation, low cost, high bandwidth along with the increased security[3]. Additionally, its been proven to better handle the privacy of both the data owner and data consumer[4]. Some blockchain based distributed file storage systems with no centralized authority are Storj[5], Sia[6], IPFS(Inter Planetary File System) & Filecoin [7-8], Swarm[9], etc. Storj and Swarm are deployed on the Ethereum blockchain network [10]. Storj provides an end-to-end encryption approach, and stores a cryptographic hash fingerprint of the file on the Ethereum blockchain while providing a method of verifying file integrity. The consensus protocol used in this system is proof of work so miners are involved in creating new blocks in the blockchain. Although proof of work is run by the successful bitcoin platform, it has its downsides of too much power consumption and minor monopoly problem. Swarm on the other hand based on distributed swarm token which is a standard ERC20 token (on the Ethereum platform) that allows one to use the Swarm software platform. It uses proof of stake protocol for incentivizing transaction in the blockchain. The SIA project forms storage contracts which are agreements between a storage provider and their client, specifying what file will be stored and at what price. The data are segmented and encrypted separately as ciphertext. The SIA coin is used as the cryptocurrency and user uploads file periodically as a proof of storage to prevent the storage node from deleting the stored file. The IPFS is more of a generic peer to peer file system with the potential to replace the HTTP protocol itself with secure file distribution with the usage of hashing each uploaded file. For incentivizing the users in getting involved in contributing to the IPFS network it also developed a native crypto coin called filecoin. Although by itself IPFS does not provide powerful cryptographic privacy protection like another peer to peer file system. Normally, in a traditional blockchain application, use asymmetric encryption with both public and private key. Usually, the private key is generated using cryptographically secured random number generator algorithm and importantly required no to be shared with anyone. From the private key, using elliptic curve multiplication public key is generated[11]. The problem is the secrecy of the private key over time. Another problem is, with a huge quantity of encrypted files, to find the required file at a reasonable time for the data consumer. Both of this issues can be tackled using Cipher-Policy Attribute-Based Encryption(CP-ABE) scheme with the secret key is related to attribute set, the ciphertext is related to specific access policy providing fine-grained access control for data consumer and also with multi-keyword search and supporting attribute revocation.[12]. The CP-ABE scheme was first proposed by Bethencourt et al. [13], but his proposed framework was not fully cryptographically secured. The system was later evolved to pass the complexity assumption of Decisional Diffie-Hellman Assumption by Li et al.[14]. Traditionally, this scheme for CP-ABE with multi keyword search is adopted by a centralized cloud storage system. Later, S. Wang et al were the first to develop such a scheme for decentralized storage systems.[15]. In addition to distributed file storage, peer to peer network is also being used in the distributed website and web application hosting.[16]. A practical example of this distributed hosting is Zeronet.[17]

II. Research outline

A. *Description*

The framework proposed here is to develop a peer to peer file distribution system with a web interface also hosted in a decentralized manner. There are two blcokchians to be maintained one is a hybrid blockchain for the structure of the distributed file system and another public blockchain for the deployment of the smart contract. One smart contract is one block on the chain will represent one service. A single smart contract will correspond to a separate blockchain for a distributed file system. A single entity can create their blockchain for providing service under a specified name. The smart contract can be created and deployed using a web interface instead of a programming language and command line interface like in the Ethereum Ecosystem. In the system, the service provider can set up the ABE scheme for data consumer, application type, etc in the smart contract. There are two types of node in the network data/service provider node and data/service consumer node that would be called a provider node and consumer node respectively. The provider node is able to create services and both the provider and consumer node plays a role in serving according to the primarily defined smart contract. Beyond this, there is no discrimination among peers in terms of network bandwidth, anonymity and/or ownership as equity is built-in to the smart contract. So, trust is embedded in the network. Getting registered to service is as easy as browsing to the distributedly hosted web application although privacy features like hiding IP address is maintained through the smart contract. Accessing the file system require registration which is secured adopting bilinear paring using elliptic curve digital signature algorithm(ECDSA). The process for registration for provider node and consumer node is defined at the smart contract level. The possible services that can be provided are public file sharing, end-to-end file sharing, file searching, upload, and download file, etc.

B. *Data collection*

Data related to the IPFS scheme and standards regarding ABE technology, blockchain underlying cryptography, consensus protocol. These are easily available on the internet as open source projects and research papers.

C. *Design*

The design decision is needed to be made regarding block metadata and consensus protocol in both the blockchain. For the smart contract blockchain, a mix of Delegated Byzantine Fault Tolerance(DBFT) and proof of concept would be used as the consensus algorithm. The distribution file system should follow the scheme of IPFS with Proof of Spacetime and Proof of Replication consensus protocol although unlike IPFS would be designed on top of the HTTP protocol. The other design works involved are in the area of smart contract web framework, registration scheme using ABE and bilinear pairing, the user interface for offered services.

D. *Implementation*

According to design, the distributed file system, smart contract blockchain, web interface for system setup and UI for the offered system is to be implemented.

E. *Analysis*

The implemented system is to be analyzed regarding the feasibility of deployment in real life scenario, performance, data availability, and integrity.

F. *Review*

After proper analysis, the system is to be reviewed to find out any design or implementation change is needed and impose those changes accordingly

G. *Report*

The report is to be prepared with the literature review, steps taking in building the system in-detail, result obtained and future scope of the research.

H. *Advice and policy formulation*

The introduction of Blockchain technology creates new challenges for both representatives at small-to-large organizations and the decision makers at the regulatory agencies. This is mainly because blockchain technology itself is as revolutionary as entering the next stage of the internet. But the open and decentralized nature of blockchains and the absence of unipolar governance means that regulatory issues need deeper analysis than previous technologies. [18] Core blockchain fundamentals like smart contracts, consensus protocols are needed to be aligned with the local and global legislation. The emergence of these newly found concepts will inevitably outdated the existing laws and regulation of the global and local governing agencies in many cases. This, in turn, means either enter new policy is to be thought and deployed or provide an adequate understanding of the newly adopted technology for the people involved (i. e. through training) without alteration of the existing laws. In deploying the proposed framework these points are needed to be noted. In addition, the framework avoided using cryptocurrency for incentivizing transactions in the blockchain for adopting the system in government agencies where cryptocurrency is banned and the person who adopts this system is discouraged to integrate digital currency into the system.

III. Research Objectives

In this section, it will discuss the key goals of the proposed research work, both the broader objectives and the specific ones.

A. Broad Objectives

The principal objective of this research is to develop a more non-centralized version of the internet that was proposed in the first place[1] by leveraging the potential of blockchain with the primary focus on a distributed file system that largely outperforms the traditional centralized system of file distribution.

B. Specific objective

- Develop a decentralized web-hosting scheme.
- Build a distributed file system on the internet following the principles of IPFS[7].
- Use blockchain to ensure secure file transfer, secure web-surfing, secure user identity.
- Content-based addressing instead of the tradition location-based addressing when it comes to surfing the internet.
- Ensure proper anonymity of the web-surfer.
- Cut cost in serving more people both in terms of finance and time by instead of buying more powerful servers, leveraging the power of a peer to peer network.
- Remove intermediation, thereby cutting cost and saving time.
- Prevent fraud on all levels.
- Apply smart contract to prevent any kind of censorship.
- Utilize non-centralization to avoid force Majeure factors i. e. natural disaster.
- Make a peer to peer system that could scale from small startups to big organizations.
- Stay aligned with the possible use cases of deploying in a government agency specifically for the people's republic of Bangladesh.

IV. Research Methodology

In this section, it would discuss the scientific methods that will be used in every process of the research.

A. Data Collection

All the data that would be collected for conducting the research are from the secondary data sources. The knowledge of the open source project that is acquired is mostly from their official whitepaper or presentation derived from the respective official website. The research papers that were consulted were are either from the "IEEE Access" the official digital library of IEEE only be accessed by an IEEE member or from the respective university website.

B. Design

The prototype for the proposed framework will be designed using proto.io[19] a fully interactive web platform for designing high-fidelity prototype. The prototype design will be based on the knowledge gathered from open-source projects i. e. IPFS[7] and research papers. Also, the scalability of the system has to be kept in mind using techniques like a smaller chain, big blocks, compressing data on the blockchain, layering in blockchain(i. e. plasma in Ethereum)[22], sharding, improving consensus protocol.

C. *Implementation*

For implementing the backend and frontend of the proposed system this research will use Google's dart[20] and dart framework flutter[21] along with the language of the web HTML, CSS, and javascript. The Integrated Development Environment(IDE) used will be Visual Studio Code.

D. *Analysis*

For analyzing the built framework a checklist is to be made that covers the minimum requirement to deploy it in an actual network. the checklist should highlight the performance, security and privacy factors. The methods adopted to test each point of the checklist should be well defined. A draft checklist is presented below:

- The system works as it was designed to work:
  - ➢ Decentralized web hosting
  - ➢ Distributed file system.
- Smart contract(s) works as it was intended
- File access time meet a certain benchmark
- Privacy criteria(i. e. IP address hidden) met
- Common web penetration testing passed(i. e. black box penetration testing, white box penetration testing)

E. *Review*

According to the result of the analysis, any alteration of the design choice and/or implementation would be considered to meet certain criteria that were not met. Again, the system would be tested against the created checklist. This would be an iterative process.

F. *Report*

Finally, the report is prepared following the IEEE standard using IEEE latex template.

V. Research Impact

The end product of the research work can put a significant impact on a large number of sectors, in Bangladesh. The great thing is the system can be deployed on an existing information distribution system without changing the current user interface and/or workflow as it is built on top of the existing internet. The training to use the new technology is minimal because most the work is being done on the backend. Despite the simplicity of its setup, it has the potential to introduce a lot of advantages over the traditional system. The system can be deployed where fraud is prevalent as the system uses blockchain technology to fully ensure user integrity. There is no single point of failure so there is no one point for malicious hackers to target. As such, it is immune to distributed denial of service(DDOS) attack which is one of the most prevailing attacks in the current world. Moreover, in the distributed file system the files are stored in multiple nodes, so even one server gets down file can be obtained from another node. The file is mapped with their hash value, the access control is protected using the digital signature with the highest cryptographic technology so only authorized users can access them. And there is no way of knowing where in the network the file is stored, makes it even more difficult for unauthorized access of data and/or tampering with the data by an outsider. Moreover, splitting files into multiple nodes results in a much higher bandwidth than centralized storage system.. All these ensure great security, high availability and data integrity for the user. In a traditional cloud storage system, it's hard to maintain the anonymity of the user. The system can ensure anonymity if privacy is a concern. Another important benefit is increased throughput but at a lower cost. For example, suppose the government will publish a board exam result that was participated by nation-wide candidates. At a time a large number of candidates will try to access the result. If in a centralized system, the existing server is not able to cope with the demand, the government needs to buy more computational power for the server and/or buy a new server to serve more persons at a time. But with the proposed peer to peer system the increased demand will actually increase the throughput and without any additional cost. Another use case of the system is to disintermediate the third parties cutting cost significantly and also provide much faster service. As an example in a government infrastructure project, a lot of third parties gets involved in supply-chain management. With the usage of smart contract in this system, the process can easily be automated with a very little workforce and in a much faster way. In short, with the correct implementation of the proposed system, it can put a huge positive influence on wherever information distribution is involved.

REFERENCES

[1] Tim Berners-Lee, “Information Management: A Proposal”, in CERN March 1989 [Online] Available: https://www.w3.org/History/1989/proposal.html

[2] Satoshi Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”, on October 31, 2008, [Online] Available: https://nakamotoinstitute.org/bitcoin/

[3] Hoang Giang Do, Wee Keong Ng, “Blockchain-Based System for Secure Data Storage with Private Keyword Search”, in 2017 IEEE World Congress on Services (SERVICES) [Online] Available: https://ieeexplore.ieee.org/document/8036727

[4] K. G. Zyskind, O. Nathan, and A. S. Pentland, “Decentralizing privacy: Using blockchain to protect personal data" in Proc. Secure. Privacy Workshops (SPW), May 2015, [Online] Available: https://enigma.co/ZNP15.pdf

[5] S. Wilkinson, T. Boshevski, J. Brandoff, and V. Buterin, ``Storj a peer-to-peer cloud storage network," [Online]. Available: https://storj.io/storj.pdf

[6] David Vorick, Luke Champine, “Sia: Simple Decentralized Storage” in Nebulous Inc., November 29, 2014 [Online] Available: https://sia.tech/sia.pdf

[7] Juan Benet, “IPFS - Content Addressed, Versioned, P2P File System (DRAFT 3)” [Online] Available: https://ipfs.io/ipfs/QmV9tSDx9UiPeWExXEeH6aoDvmihvx6jD5eLb4jbTaKGps

[8] “Filecoin: A Decentralized Storage Network” in Protocol Labs, July 19, 2017 [Online] Available: https://filecoin.io/filecoin.pdf

[9] Philipp Pieper, Timo Lehes, “SWARM FUND White Paper The Blockchain for Private Equity” in March 2018 [Online] Available: https://docs.swarm.fund/swarm-whitepaper-eng.pdf

[10] G. Wood, “Ethereum: A secure decentralizedgeneralized transaction Ledger”, [Online] Available: https://gavwood.com/paper.pdf

[11] Bos, Joppe & Alex Halderman, J & Heninger, Nadia & Moore, Jonathan & Naehrig, Michael & Wustrow, Eric. (2014). Elliptic Curve Cryptography in Practice. 8437. 10.1007/978-3-662-45472-5_11. [Online] Available: https://www.researchgate.net/publication/274651795_Elliptic_Curve_Cryptography_in_Practice

[12] Wang S, Yao L, Zhang Y (2018) Attribute-based encryption scheme with multi-keyword search and supporting attribute revocation in cloud storage. PLoS ONE 13(10): e0205675. https://doi.org/10.1371/journal.pone.0205675 [Online] Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205675

[13] Bethencourt J, Sahai A, Waters B. “Ciphertext-Policy Attribute-Based Encryption” in 2007 [Online] Available: https://ieeexplore.ieee.org/document/4223236

[14] Jiguo Li, Haiping Wang, Yichen Zhang and Jian Shen, "Ciphertext-Policy Attribute-Based Encryption with Hidden Access Policy and Testing," KSII Transactions on Internet and Information Systems, vol. 10, no. 7, pp. 3339-3352, 2016. DOI: 10.3837/tiis.2016.07.026 [Online] Available: http://www.itiis.org/digital-library/manuscript/1408

[15] Shangping Wang, Yinglong Zhang , Yaling Zhang, “A Blockchain-Based Framework for Data Sharing With Fine-Grained Access Control in Decentralized Storage Systems” [Online] Available: https://ieeexplore.ieee.org/document/8400511

[16] Svebor Prstačić ; Mario Žagar, “A model for web application and web service peer-to-peer hosting network architecture” in 2013 [Online] Available:

[17] “ZeroNet: Decentralized web platform using Bitcoin cryptography and BitTorrent network” Presentation, [Online] Available: https://zeronet.io/files/ZeroNet_Presentation.pdf

[18] Jason Potts, “Public Policy for Cryptocurrency and Blockchain Technology”, [Online] Available: www.ippapublicpolicy.org/panel/pdfPanel.php?panel=225&conference=7

[19] “Proto.io - Prototypes that feel real”, https://proto.io/, last accessed: 10:55 AM 6/9/2019

[20] “Dart programming language | Dart”, https://dart.dev, last accessed: 12:26 PM 6/9/2019

[21] “Flutter - Beautiful native apps in record time”, https://flutter.dev, last accessed: 12:26 PM 6/9/2019

[22] “Understanding Plasma | Ethereum Scaling | State Channels vs. Plasma”,https://education.district0x.io/general-topics/understanding-ethereum/understanding-plasma/ last accessed: 2:38 PM 6/9/2019