International Journal of Internet, Broadcasting and Communication Vol.13 No.1 26-36 (2021)
http://dx.doi.org/10.7236/IJIBC.2021.13.1.26

IJIBC 21-1-4

Blockchain for the Trustworthy Decentralized Web Architecture

Geun-Hyung Kim

Professor, Game Engineering Major, Dong-eui University, Korea
geunkim@deu.ac.kr

Abstract

The Internet was created as a decentralized and autonomous system of interconnected computer networks used for data exchange across mutually trusted participants. The element technologies on the Internet, such as inter-domain and intra-domain routing and DNS, operated in a distributed manner. With the development of the Web, the Web has become indispensable in daily life. The existing web applications allow us to form online communities, generate private information, access big data, shop online, pay bills, post photos or videos, and even order groceries. This is what has led to centralization of the Web. This centralization is now controlled by the giant social media platforms that provide it as a service, but the original Internet was not like this. These giant companies realized that the decentralized network's huge value involves gathering, organizing, and monetizing information through centralized web applications. The centralized Web applications have heralded some major issues, which will likely worsen shortly. This study focuses on these problems and investigates blockchain's potentials for decentralized web architecture capable of improving conventional web services' critical features, including autonomous, robust, and secure decentralized processing and traceable trustworthiness in tamper-proof transactions. Finally, we review the decentralized web architecture that circumvents the main Internet gatekeepers and controls our data back from the giant social media companies.

Keywords: *Blockchain, decentralized web, future internet service architecture, decentralized Internet, stateful web, trustworthy decentralized web.*

1. Introduction

Haber and Stornetta envisioned many concepts of blockchain technology [1]. Their work focused on timestamping documents; this process verifies that a document was created at a specific time in a specific version by storing hash values in a timestamped block on a tamper-proof blockchain. They also adopted the Merkle tree to enhance timestamping efficiency by enabling a single block to include many documents. Satoshi Nakamoto, in 2008, conceptualized the first peer-to-peer version of cryptocurrency using the blockchain technology and described how the blockchain technology was well equipped to strengthen digital trust in terms of decentralization; these blockchains did not require trusted intermediaries [2]. Many IT innovators and

Manuscript Received: November. 5, 2020 / Revised: November. 11, 2020 / Accepted: November. 15, 2020
Corresponding Author: geunkim@deu.ac.kr
Tel: +82-51-890-2271, Fax: +82-0505-182-6910
Professor, Game Engineering Major, Dong-eui University, Korea

experts regard blockchain technology as one of the most meaningful technological inventions in recent years, digitizing secure ownership of assets. The technology is based on the concept of a distributed ledger, decentralized cataloging, and large data description. Many consider also blockchain as a disruptive technology that will revolutionize business and redefine companies and economies.

The Internet was originally invented as a decentralized autonomous system in which a participant communicates to each other using peer-to-peer interconnectivity without relying on a single server. However, the Web's advent, especially Web 2.0 [3], allowed users to generate their own data, share them, collaborate, and utilize service-side scripting to proliferate online services based on user's data. Giant social media platforms have built value using free-obtainable private, personal data that have been deployed on the Web. Therefore, the models for applications and several service infrastructures on the Web (Internet) have become more centralized differently from the original architectural intentions due to the business models that depend on centralized accounting and administration [4].

The Web had been one of the representative open application platforms on the Internet since Tim Berners-Lee invented it approximately 30 years ago. It connects significant amounts of information on the Internet in a hypertext manner, providing users a platform to search for the same conveniently. The Web technology basically consists of three main components: the URL (unified resource locator) representing the location to a specific site; HTTP (hypertext transfer protocol), which is the protocol for sending and receiving request and response messages; and HTML (hypertext markup language), which is the markup language for creating hypertext pages easily. Web 1.0 and Web 2.0 have revolutionized information and interactions, respectively. Presently, relatively few social media platforms are responsible for hosting essential elements of what we consider the Internet and keeping our email, social media, and web pages available to all. These giant social media platform providers own hosting servers and exercise control over how the Internet operates. The current centralized Web platform exposes the Internet to certain vulnerabilities, which are likely to worsen shortly. These vulnerabilities are mainly related to scalability and availability of resources and services, reliability for a single point of failure and bottleneck, security and privacy for user data, and trust-ability [5-6]. In this study, we focus on these issues and provide a review of the potentials and capabilities of blockchain for decentralized web architecture.

The rest of the paper is organized as follows: In Section 2, we present the related work, and we discuss the decentralized web architecture in Section 3. Finally, we present the conclusion of our study and discuss future works and trends.

2. Related work

2.1 Issues in Centralized Web

The Internet has been acting as a digital information highway that we can use ubiquitously today. The web has been the enabler for the Internet to become a hub to exchange information. The Web was initially designed in a decentralized fashion since the information on the Web can be linked to the information stored on other computers on the Internet. In the first phase (called Web 1.0) of the Web's evolution, content creators were a few and the majority of users only acted as consumers of contents [7]. The open Web platform, i.e., the collection of open technologies enabling the Web, emerged by the early 1990s; it has driven the Web 2.0 era. Any participant in Web 2.0 can be a content creator owing to the emergence of newer technologies, such as

mashups, AJAX (autonomous JavaScript and XML), and REST API (representational state transfer application programming interface) in the open Web platform. The essential characteristics of Web 2.0 are openness, freedom, and collective intelligence by way of user participation [7]. With the advent of Web 2.0, users began to interact and collaborate among themselves and share information via centralized social media platforms provided by established companies. After a long time of focus on the front-end of the web (Web 2.0), the evolution of web utilization and interaction among several paths have enabled the upgrade of the back-end of the Web. The Semantic Web (called Web 3.0) necessitates using a declarative ontological language like OWL (web ontology Language) to produce domain-specific ontologies that machines can infer information and make new conclusions, not simply match keywords [8]. The comparison of Web 1.0, Web 2.0, and Web 3.0 is shown in the Table 1.

Table 1. The comparison of Web 1.0, Web 2.0, and Web 3.0 [8]

Features	Web 1.0	Web 2.0	Web 3.0
Data access	mostly Read-Only	Read-Write	Portable and Personal
Content usage	Owning	Sharing	Consolidating
Service level	Web form	Web application	Smart application
Info. contact point	Directory	Tagging	User behavior
Info. flow	Uni-directional flow	Bi-directional flow	Multi-directional flow
Formats	HTML	XML / RSS	RDF / RDFS /OWL

As the commercial prospects of the Web grew along with the development of Web technologies, many service platforms related to social media emerged. Presently, only a few mighty platform companies control most of these social media platforms on the Internet. These giant platform companies are popularly, jointly known as FAANGs (Facebook, Amazon, Apple, Netflix, Google, Microsoft, and Twitter). Over time, the Web has become more centralized in terms of its architecture and technology. In the centralized Web, social media platform providers have monopolized control over user data. The data monopoly is giving rise to a new set of problems alongside our online communication consolidation. The data monopoly may cause the biggest issue that social media platform providers collect user data and sell it to an interested third party.

With all the user data in the hands of a few monopolistic companies, users are becoming increasingly vulnerable to hacking, surveillance, censorship, data breaches, misinformation, and so on [6]. For example, at the behest of the Chinese authorities, Google's decision to operate a censored search engine in China has raised concerns among human rights groups about the future of the Web [10]. The Egyptian government had blocked around 500 websites as of February 2018. The Internet traffic to and from Egypt across 80 internet service providers worldwide dropped precipitously on January 27–28, 2011. In 2011, thousands of websites experienced downtime because of the Amazon Web Services (AWS). The cause of these massive Internet outages is the monopolistic nature of major cloud service providers [6]. With the emergence of centralized web service platforms, these select companies have control over vast user data. Therefore, Web users are losing control over the content they read. These problems have also compromised Web users' privacy, which may

make them easy targets for hackers.

Nearly all demographic groups in the US consider social media the most dominant news source. In fact, Facebook is the most commonly used source of news on government and politics for Millennials, especially. Thus, a few platforms can significantly influence what media users consume daily. Hence, these platforms can control what is possible to publish and what content is likely to be discovered [10]. The following are the risks posed by the centralized Web [6]. Since its main characteristic is to guarantee the safety and integrity of data on a decentralized network, blockchain can be applied to solve the following risks; direct censorship, indirect censorship, abuse of curatorial power, abuse of privacy, monetize data, exclusion, single point of failure, data breach, and so on.

2.2 Decentralized Storage Platform

A decentralized storage platform would be considered better than a centralized storage platform to address scalability and trust issues in data storage and sharing. To solve a single-point failure, the decentralized storage platform ensures that no single server controls the entire service. User's data is distributed and stored in peer-to-peer interconnected storage nodes that constitute a decentralized storage platform. The peer-to-peer network can operate smoothly so that users can access the data at any time. In the decentralized storage platform, the user retains ownership of his/her data and be the only one to access his/her data or permit other users [11]. Decentralized Storage Platform consists of data integrity verification, data encryption, data decentralization functionality with blockchain shown in Figure 1.

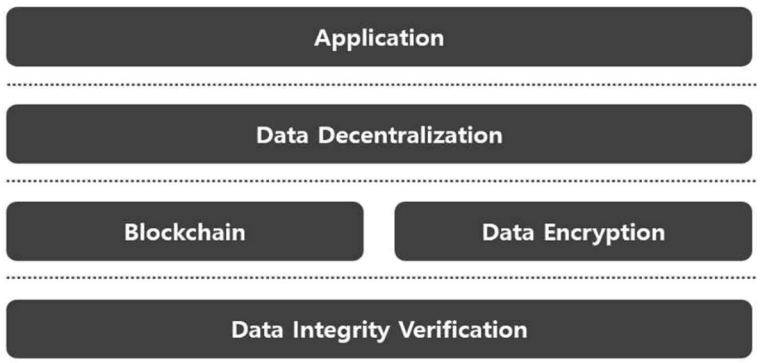

Figure 1. Functional layers in Decentralized Storage Architecture [11].

2.3 Blockchain as a Value Highway

Blockchain [2] is an immutable digital ledger that records and verifies cryptographically signed transactions grouped into blocks in a distributed fashion without a central authority. Except for the genesis block in a blockchain, each block cryptographically points to its immediately previous one after undergoing a distributed consensus decision and validation. The blockchain platforms maintain the blocks containing the electronic cryptographic data in a distributed and consensus way. Based on a distributed blockchain and consensus-based maintenance, individually developed policing mechanisms ensure that valid transactions are added to the blockchain that allows users to be pseudonymous or anonymous; users can create accounts without identifying the authorization process. Therefore, applications built on the blockchain can enable the business to be with untrusted and unknown users.

Blockchain is based on decentralized peer-to-peer networking, in which all participant nodes provide their

resources fairly, alleviating one-to-many traffic flow bottlenecks. The characteristics of blockchain are to provide decentralization, transparency, non-repudiation, and traceability [12]. A blockchain is a public registry of who owns and who transacts what. The transactions are secured through cryptography, and the transaction history gets locked in blocks of data that are then cryptographically linked together and secured. This creates an immutable record of all the transactions across this network that cannot be forged. The record is replicated on every node in the network. Unlike the existing Internet, blockchain enables users to deliver value without relying on a third party. As of now, it has been used to deliver various economic values such as cryptocurrency, stocks, computing resources, real estate, automobile use rights in a shared economic society, and intellectual property rights.

Several cryptographic technologies, such as hashing and digital signature, have been used in blockchains. Hashing is a method of calculating a relative unique fixed-size output (called digest) for the input of nearly any size (e.g., a video stream, a text file, or an image) and is designed to be one-way and collision-free. Because it results in completely different digests, even if a single bit in the input data is modified, it provides the integrity of a block data in the blockchain. For digital signatures, asymmetric-key cryptography is utilized; this provides the ability to verify someone's identity who participates in a transaction. Each user possesses a pair of private and public keys. The private key, regarded as the user's identity and security credential, is used to sign transactions digitally; the digitally signed transactions are sent to whole nodes. The public key is used to validate the transactions that are signed with the private key. When a new transaction occurs, the user submits a new transaction to the blockchain ledger. The new transaction will be copied and distributed among every node in the blockchain platform. It will be stored in a queue until a mining node adds it to the blockchain by creating a block.

Blockchains allow us to write code and have binding contracts between individuals and then guarantee that these contracts will be enforced without a third party's requirement. Blockchains redefine how digital trust mechanisms work using distributed consensus mechanisms and transparent tamper-evident record-keeping.

3. Trustworthy Decentralized Web Architecture

Current Web applications combine service and data and execute with a closed back-end database. Because of this coupling, an RSVP on an application event will not be reflected in the scheduler. Thus, similar or related data will be redundantly stored in multiple applications, such as the centralized Web application shown on the left side in Figure 2. The applications in the centralized Web compete in a single market based on data ownership. New innovative competitors, in the centralized Web ecosystem may struggle to enter market because of a lack of customer data. In this section, we introduce the trustworthy decentralized Web architecture and the data model and functional components for identity and data management.

Fundamentally, decentralizing the Web is about enabling choice by breaking up artificially coupled decisions into individual options that can be combined at one’s pleasure [13,14]. In a decentralized Web, we should be able to interact with websites and other people without commitment to a single social media platform. Sensitive personal data should be decoupled from applications in terms of taking back control of it. This separation allows users to enjoy the applications they want and store data where they specify. Also, this allows service providers to develop applications without the accumulation of their own user data. An example of the data and service separation for a decentralized app application is shown on the right side in Figure 2 [13]. End-

users can select any service provider to store their text, photos, and videos on their own storages on the Internet and depend on any third-party services to interact with data, regardless of storage location. As an example, identity data for a crucial identity service can be provided by Web storage.

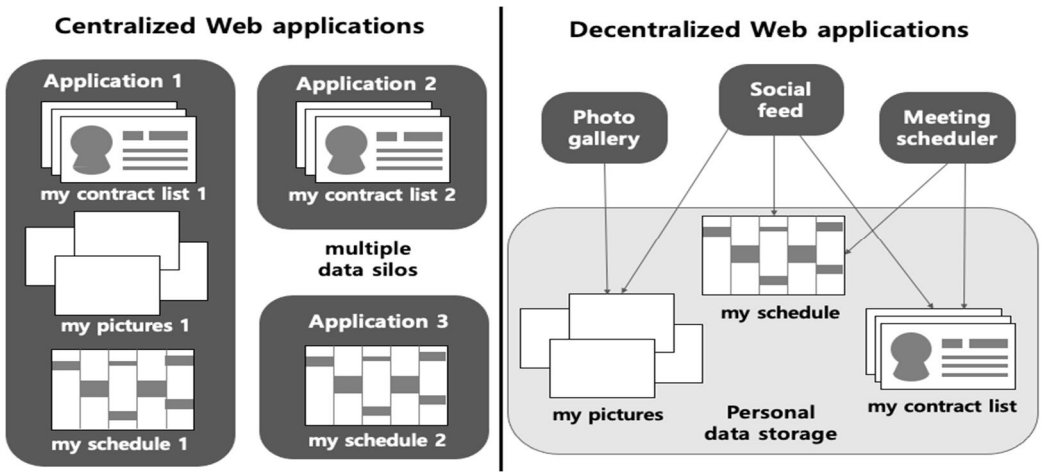

Figure 2. Comparison of centralized Web and decentralized Web.

In the decentralized web ecosystem, end-users have the right to control their data, unlike the centralized web ecosystem. However, although users can control the data in the decentralized web ecosystem when the service providers utilize the user's data, they can infer the data owner and infringe the user's privacy. Besides, as data generated by wearable devices and sensors in the home network and user-generated contents become valuable, data sovereignty is becoming more critical. Hence, it is desirable that the identifier be anonymized or pseudonymized to ensure the user's privacy, and if possible, a relationship between both identifiers of data and owner must be established to claim data ownership.

3.1 Identity Management

This study proposes the self-sovereign IDMS (identity management system) for trustworthy decentralized Web architecture to anonymize or pseudonymize identity and mitigate data and identifier's coupling. Proposed self-sovereign IDMS is based on the self-sovereign identity (SSI) abstract architecture that consists of four components; decentralized identifier (DID), decentralized key management system (DKMS), DID auth, and verifiable credentials (VCs).

A DID is a new type of identifier that is unique worldwide, verifiable with high availability, and verified cryptographically. In contrast to the traditional and federated identifier, the DID was designed to decouple from centralized registries, identity providers, and certificate authorities. Figure 3 shows the self-sovereign identity management model deployed in the proposed architecture and can be implemented based on the DID concept. Because the DID can identify any subject like people, organizations, groups, digital objects, and abstract entities, we adopt the DID to identify digital data, logical/physical objects, and owners in the trustworthy decentralized Web ecosystem.

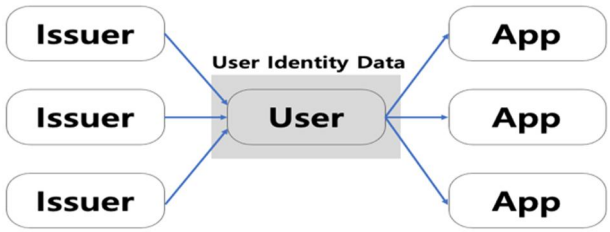

Figure 3. Self-Sovereign Identity Management Model.

We design the self-sovereign identity management architecture that uses DIDs and pairwise VC and the zero-knowledge proofs. The VC is a portable and provable claim about a subject and an interchangeable and cryptographically verifiable digital certificate. It refers to the subject's DID and includes the issuer's DID. The relationship between DIDs and VCs is depicted in Figure 4. A DID Document (DDo) with that a DID is associated, includes the DID itself, public key to authenticate and encrypt messages, service endpoint. An example of a person's VC could be the person's name, social security number, etc. The physical/logical things may claim to be a type of sensor and the data file has an owner's DID as a VC's entity. To ensure privacy, actual VCs, DIDs and private keys are stored in a personal digital wallet, hosted by user's mobile devices and shared if necessary. In the proposed architecture, the algorithm for zero-knowledge proofs is applied not to disclose the actual value when the verifier verifies the credentials.

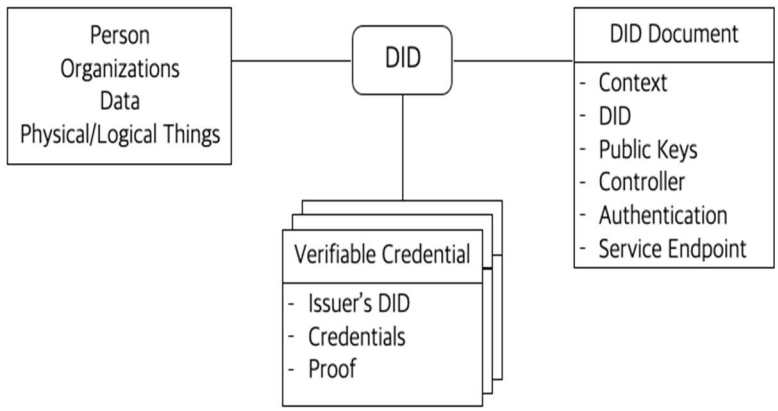

Figure 4. The relationship between DIDs and VCs.

The ownership represented with DID document and VC is depicted in Figure 5. All subjects in the proposed architecture generate DID documents. The owner of data issues the VC to represent the owner. The DID documents of the data owner and the data file include its DID, respectively, and the VC issued by data owner contains the issuer's DID, the credential representing the ownership, and the proof.

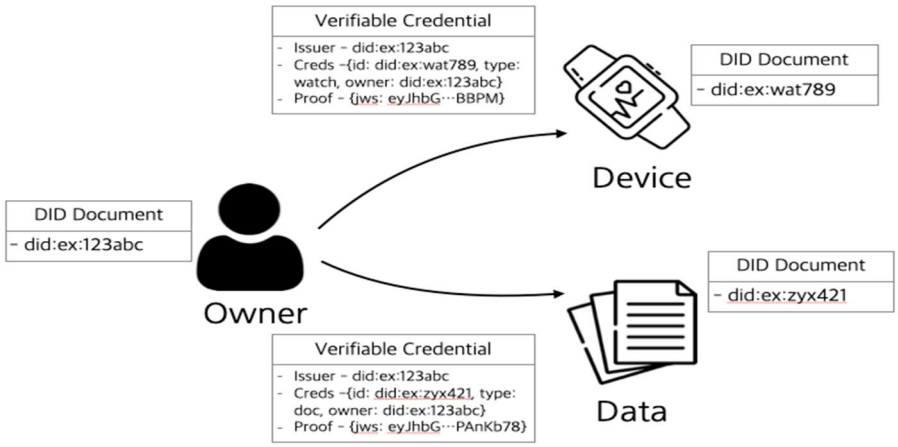

Figure 5. The DID Document and VC representing the ownership.

The identity data manager in the proposed architecture is designed to issue DIDs and VCs for owners, devices, and data, sign data transactions, request the user's connection endpoints, and validate the signers. With DID/VC-based user, device, service authentication, we verify if the subjects have the rights to utilize the verified service, device, or data in the proposed architecture. Besides, we verify if the transaction is coming from a verified subject in the architecture with DID/VC-based transaction verification. DIDs and VCs are interoperable across local services and other decentralized identity systems and follow the standards of world wide web consortium (W3C). Consequently, DID/VC-based identity management mechanism proves the possession of identifier and attributes without a central authority.

3.2 Data Management

The proposed architecture enables a decentralized web ecosystem in which various decentralized applications, such as file sharing, IoT applications, and more enterprise-driven applications like contract signing and transaction or service level monitoring. These applications utilize data management function to store data distribution and data ownership related to their user's authorization. The front-end components of the Web application are stored in the decentralized manner discussed in the previous section about decentralized storage platform.

Data management components in the proposed architecture manage data generated by applications using the decentralized ledger and the decentralized storage platform. The decentralized Web applications request to store data to the data management component that store actual data to the decentralized storage platform and data transactions to the decentralized ledger. The proposed architecture does not utilize central data servers and provides secure peer-to-peer data and transaction exchange with good performance. Because the decentralized storage platform manages the data distribution across the distributed network by storing file objects and distributed hash table (DHT), it provides resiliency without Internet access keeping the network alive in a node failure or a node disconnection. In the proposed architecture, the binding mechanism between the transaction on the decentralized ledger and actual data on the decentralized storage platform is required to prohibit the direct disclosure of personal data and protect privacy. Binding between the transaction on the decentralized ledger and the actual data is accomplished by storing the pointers to the actual data on the decentralized storage platform. The pointers act as proof that transactions are taking place and are stored on a

distributed ledger and accessible by all nodes.

The decentralized ledgers for identity and data perform identity and data management, respectively. Using the decentralized ledgers for data is to store transactions and pointers to actual data stored on the decentralized storage platform. The purpose of using the decentralized ledgers for identity is to support self-sovereign identities and verifiable credentials. This paper proposes the concept self-sovereign data by extending the decentralized identifier to express data ownership in the decentralized Web architecture. Besides, we propose an integrated decentralized ledger that combines identity management and data management. Figure 6 shows the self-sovereign identity and data management architecture.

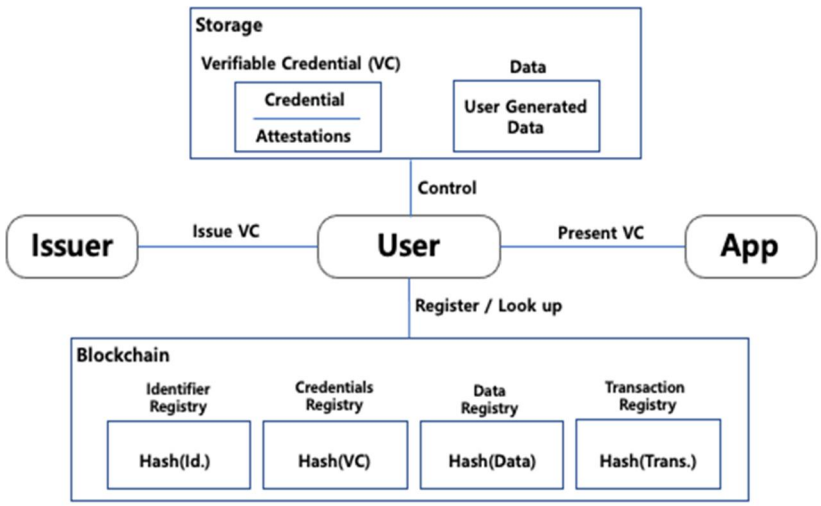

Figure 6. Self-Sovereign Identity and Data Management Architecture.

Blockchain in self-sovereign identity and data management architecture in Figure 6 will store the hash values of verifiable credentials, user-generated data, identifiers, and data transactions (e.g., reading/writing data), respectively. The separation of data, services, and identifiers in the proposed architecture is achieved by blockchain technology. Consequently, it is feasible to build a decentralized Web ecosystem in which individual self-sovereignty is secured with the proposed architecture.

Today, the Internet does not have a mechanism to transfer the status of identity, ownership, authentication, and authorization. However, the state is a crucial property for managing values. Blockchain technology has introduced a way for each participant in the network to hold and transfer value in a digitally native format without trusted central intermediaries. The consensus protocol is designed in a manner that the blockchain can collectively remember preceding events or user interactions. Blockchain protocol can be considered as a game-changer, opening the way to a more decentralized Web ecosystem.

4. Discussion

We propose the trustworthy self-sovereign decentralized Web architecture based on DIDs and VCs representing ownership among subjects. The proposed architecture contains decentralized ledgers for transactions and identities, respectively, along with a decentralized storage platform. We designed a data management component for transparent data exchanges and an identity management component for

guaranteeing user privacy and data sovereignty.

5. Conclusion

Today, increasingly user-generated content is hosted on web servers belonging to a small group of giant companies. This trend has created the centralized Web with several issues, censorship, surveillance, abuse of curatorial power, abuse of privacy, monetize data, data breach, not the least of which is giant companies holding power through vast amounts of data. Decentralized Web has the potential to revolutionize contracts and value exchanges and to decouple data and related applications. The decentralized web basically should separate data and applications, give data rights to data producers, and ensure responsible data sharing. We discuss self-sovereignty identity, self-sovereignty data, and trustworthiness for data or value transactions that are a necessity in the trustworthy decentralized web. Therefore, we discuss self-sovereign identity, self-sovereign data, and trustworthiness for data or value transactions necessary in the trustworthy decentralized web. In this study, we reviewed the issues related to centralized web architecture, decentralized storage platform, and blockchain technology and discussed blockchain's potential for decentralized web architecture. Subsequently, we discussed a trustworthy decentralized web architecture that circumvents internet gatekeepers and takes control of our data back. In the future, we will implement the designed trustworthy decentralized Web architecture using open source blockchain technology and a decentralize storage platform.

Acknowledgement

This work was supported by Dong-eui University(20200362001) and National Research Foundation of Korea (NRF) grant founded by the Korea government (MOE: Ministry of education): (NRF-2017R1D1A1B03035074)

References

[1] S. Haber and W. S. Stornetta, "How to time-stamp a digital document," *Journal of Cryptology*, Vol. 3, No. 2, pp. 99-111, Jan. 1991. DOI: https://doi.org/10.1007/BF00196791

[2] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. *https://bitcoin.org/bitcoin.pdf*.

[3] S. Murugesan, "Understanding Web 2.0," *IT Professional*, Vol. 9, No. 4, pp. 34-41, July-Aug. 2007. DOI: https://doi.org/10.1109/MITP.2007.28

[4] Charter for Research Group – Decentralized Internet Infrastructure Research Group (DINRG), 2020. *https://datatracker.ietf.org/group/dingr/about*.

[5] W. Yang, E. Aghasian, S. Garg, D. Herbert, L. Disiuta, and B. Kang, "A Survey on Blockchain-Based Internet Service Architecture: Requirements, Challenges, Trends, and Future," *IEEE Access*, Vol. 7, pp. 75845-75872, May 2019. DOI: https://doi.org/10.1109/ACCESS.2019.2917562

[6] G-H. Kim, "How will blockchain technology affect the future of the Internet?," *in Proc.14th CUTE 2019, pp.1-6,* Dec. 18-29, 2019.

[7] G. Cormode, B. Krishnamurthy, "Key differences between Web 1.0 and Web 2.0," *First Monday*, Vol. 13, No. 6, June 2008. DOI: https://doi.org/10.5210/fm.v13i6.2125

[8] GeeksforGeeks, Web 1.0, Web 2.0 and Web 3.0 with their difference, *https://www.geeksforgeeks.org/web-1-0-web-2-0-and-web-3-0-with-their-difference/*.

[9] Z. Corbyn, Decentralisation: The next big step for the world wide web, Article of the Guardian, *https://www.theguardian.com/technology/2018/sep/08/decentralisation-next-big-step-for-the-world-wide-web-dweb-data-internet-censorship-brewster-kahle*.

[10] C. Barabas, N. Narula, and E. Zuckrman, "Defending Internet Freedom through Decentralization: Back to the Futures," *report of MIT digital currency initiative and the center for civic media,* 2017.

[11] Y.-J. Han, G-H. Kim, "A Comparative Study on Decentralized Storage Platform for Self-sovereign Data," *Asia-pacific Journal of Convergent Research Interchange*, Vol. 6, No. 5, pp. 1-10, May 2020. DOI: https://doi.org/10.21742/apjcri.2020.05.01

[12] H.-N. Dai, Z. Zheng, Y. Zhang, "Blockchain for Internet of Things: A Survey," *IEEE Internet of Things Journal*, Vol. 6, Isue 5, pp. 8076-8094, Oct. 2019. DOI: https://doi.org/10.1109/JIOT.2019.2920987

[13] R. Verborgh, Re-decentralizing the Web, for good time, *https://ruben.verborgh.org/articles/redecentralizing-the-web/.*

[14] S. Voshmgir, *Token Economy: How the Web3 reinvents the Internet*, BlockchainHub Berlin, pp. 27~34, 2020.