Research on an Intelligent and Intrinsic Security Data Circulation System Based on the Infrastructural Enablers for Web 3.0

Qingqing Tu(✉)

National Computer Network Emergency Response Technical Team/Coordination Center Sichuan Sub-Center, Beijing, China
tuqingqing@cert.org.cn

Abstract. With the development of digital technologies, data security is under various risks, which impedes the data circulation to release the data value. Fortunately, the development of Web 3.0 provides a novel paradigm to solve this problem. With recent advances, Web 3.0 is considered a ubiquitous and decentralized value-sharing network to drastically revolutionize our life and society. By analyzing the characteristics of data circulation and based on the key categories of infrastructural enablers for Web 3.0, our paper proposes an intelligent and intrinsic security data circulation system for the next-generation internet. The proposed system enhances the trust in data interaction and stimulates the deep data circulation from six aspects, including hierarchical management, intelligent circulation, application support, identity management, security protection, and evaluation and supervision, which also provides a new way to cope with the new challenges in the Web 3.0 data ecosystem.

Keywords: Intrinsic security · Web 3.0 · Intelligent computation · Data circulation

1 Introduction

With the improvement of technology, the Internet has dynamically evolved from the initial "read-only" Web 1.0 period to the "readable, writable and participable" Web 3.0 era. Facing the decline of Internet traffic dividend and the increase of data security risk caused by the concentration of large platforms, Web 3.0, integrated with multiple technologies, builds a decentralized, secure, and trusted Internet of value, and helps transfer the value to individuals, by breaking the ecological boundaries of the traditional Internet [1], which is a new direction of Internet iteration. As the digital economy comes to the network-based period, data becomes the key factor connecting the online virtual space and offline real spaces, and the strategic resource to promote technological innovation and value mining in the digital economic system.

G. Guan et al. (Eds.): ICBDSS 2022, AHCS 8, pp. 50–60, 2023.
https://doi.org/10.2991/978-94-6463-064-0_7

However, more data does not necessarily mean more data value. As an important way for data to create value, data circulation can help data integrate with the production and life and tap demand to release value and create opportunities. Until recently, the digital development of most cities is still in the primary stage, and data is discretely distributed. Since the data subjects with different scales of data and processing capacity, there are still a considerable number of the data subjects unclear about their own data resources. What's more, the problem of data security and quality impedes the smooth data circulation and cooperation between subjects.

With the help of multi-party computing, artificial intelligence, and blockchain technology, data circulation is evolving from traditional "single-to-single" to "multi-to-multi". Since the data with the feature of multi-subjectivity, reuse, and portability, the increasing number of data subjects and the extended data processing activities lead to the paths and methods of data circulation being more complex and dynamic. Therefore, an increasing risk occurs in the data circulation life cycle. Specially, when external attack threats continually increase, data security incidents occur frequently. According to the security monitoring result in Fig. 1, with the increase in data scale and data importance, the security comes more severe. There were 4,145 data breaches already publicly disclosed worldwide.

Therefore, both data security and data circulation are key factors to promote the development of the digital economy and the important components of the future Web 3.0 ecosystem. However, there is still less research on security data circulation systems based on the advantages of Web 3.0, which motivates our work. In this paper, we first analyze the characteristics of data circulation and the key categories of infrastructural enablers for Web 3.0, then we propose a novel way to construct an intrinsic security data circulation system for the next-generation internet. Based on the privacy enhancement technology the proposed system enhances the trust for data interaction and stimulates deep data to circulate, which also provides a new way to cope with the new challenges in the Web 3.0 data ecosystem.

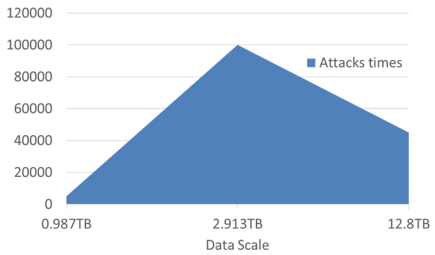

Fig. 1. Attack times for the data subjects under different data scale.

2 Related Work

Numerous works tried to define Web 3.0 in the past few years. With the evolution of blockchain infrastructures and the development of decentralized applications, recently, there is a work that gave out the fundamental and the first academic definition for Web 3.0 [2]. It contains three enable factors, including (1) security and performance-enhanced platforms to support data computing based on individual blockchains; (2) the federated or centralized platforms which can publish verifiable states; (3) a secure interoperability platform to connect these distributed data subjects. In this way, Web 3.0 provides a unified and connected computing platform for value-sharing realization. Although the application of Web 3.0 is unclear, security and data circulation are the infrastructural enablers for Web 3.0, which are also the advantages of Web 3.0 based on the techniques to realize it.

Since the characteristics of data include multi-subject, portability, reusability, and liquidity, it is expecting a novel secure, and efficient circulation mode. The privacy enhancement technology provides a possibility for the construction of a secure data circulation system. To improve computing power, transmission efficiency, and security in data circulation, the researchers are mainly in two ways: (1) one way is based on Trusted Execution Environment (TEE), which is a private computing scheme based on hardware and cryptography principles. There are many research and application that realizes the TEE technique by building an encrypted Environment isolated from applications and system, such as Intel SGX, TrustZone of ARM, TICS of HUAWEI, etc. The disadvantage of this method is the requirement of a Trusted Third Party (TTP). (2) The other way is through Privacy-Preserving Computation, which allows multiple data subjects to cooperatively compute the specified objective functions. Besides the calculation results, the input information of other data subjects cannot be obtained during the calculation process. There are two main types of technology: first is the secure multi-party Computation (MPC) based on Secret Sharing (SS), Garbled Circuit (GC) algorithms, or other cryptography algorithms [3], which have already been applied to Alibaba's Ant chain MORSE security computing platform, Tencent Angel PowerFL federal computing platform, Huafang Qingjiao multi-party security computing platform, etc. The second is Federated Learning (FL) technology that integrates artificial intelligence and privacy protection technology for joint modeling, training, and computing [4], which is also applied in the Federated Computing Service platform of China Mobile, Ali Cloud Machine learning PAI, Tianyi Cloud Zhuge AI-Federated Learning platform, etc. What's more, there also have several specific privacy protection technologies, including Homomorphic Encryption for ciphertext computation [5], Differential Privacy (DP) that confuses private data to avoid direct identification [6], Zero-knowledge Proof (ZKP) to avoid identity information leakage [7], etc. The comparison of different privacy-enhancing technologies is shown in Table 1. Privacy enhancement technique has been gradually from a single algorithm to hybrid technologies development, and from simple to complex applications calculation. In the current existing application scenarios, a variety of privacy enhancement methods are combined to use to complete the efficient data calculation, analysis, and model construction task, which is not at the expense of side effects on data security and personal privacy. However, privacy-enhancing computing is still in the primary stage of development, and still faces challenges including security, complexity,

Table 1. Privacy Enhancing Technology Analysis

Categorization	TEE	Privacy-Preserving Computation	
		MPC	FL
Secure technology	Trusted hardware environment	Cryptography based SS and GC technologies, etc.	Joint modeling techniques combining machine learning and privacy protection techniques
Controllability	middle	high	high
Performance and complexity	high	low	middle
Privacy Protection Capability	middle	high	middle
TTP	Need	No	No

and cost. Specially, when data circulation between heterogeneous systems, there exists a barrier for data circulation. Fortunately, some studies [8] by introducing intelligent middleware and blockchain smart contract solved the problem of data interaction, and the underlying communication, and achieved a relatively low coupling easy to expand the train of thought, However, relevant problems need to be further solved and the scheme needs to be further optimized.

As mentioned above, Web 3.0 builds a new ecology of co-construction and sharing based on cryptography and decentralized technology, and it complements the advantages of privacy enhancement technology to build a more secure and efficient data circulation way.

3 The Features Analysis of Data Circulation in Web 3.0

Web 3.0 aims to create the Internet of Value, which has great differences in development path and the technologies with the traditional Internet, including the blockchain as the underlying technology to break the boundaries of the traditional platform, any data subject can profit from the data and create value on their own, and the integration of online and offline data lines leads to more frequent value exchange. Therefore, as an important part of Web 3.0, data circulation presents characteristics different from the traditional Internet era to adapt to the development of Web 3.0.

3.1 The Dispersed Data Restrain Data Circulation

Web 3.0 promotes enhancing the autonomy of the individual based on decentralized technologies. In the past, data were collected, processed, and earned by the Internet platforms, but now any data subject can achieve the benefits brought by the value of data. Thus, valuable data will be widely distributed to discrete data subjects. According to the

"economies of scale", the decentralization in the Web 3.0 era leads to data fragmentation and distribution, the full release of data value requires more data subjects to participate in the circulation. However, data mining is a professional and difficult work with technical requirements. When valuable data return to small enterprises or even individuals, it may be difficult to use due to the costs and ability problems, which lead to isolated data islands and affect the data value mining.

3.2 Massive Data Require to Balance Security and Efficiency

With the advent of the Internet of Everything, data scale will increase by thousands of times [9]. Although technology integration provide new possibilities for improving data security protection and computation ability, mapping the real world in the virtual world produces more important and sensitive data. Therefore, the attack led to severe damage. As mentioned previously, the life cycle of data circulation involves multiple periods, such as collection, storage, processing, and transmission. And risks are concurrent at multiple periods, which increases the difficulty of risk control. In addition, with the increase of data interaction subjects, there will be increasing pressure on security attacks, such as DDOS attacks, and reconstruction attacks launched by illegal subjects in federated learning [10], which endanger data security. However, the traditional data protection schemes often apply to a single system or reduce data availability. In Web 3.0, it is necessary to consider the whole life cycle of data circulation synchronously to provide effective protection for data processing. While the new scheme should also consider the different scales of the data subject, the operation ability, and the security input capacity.

3.3 Data Value Require the Data Control

Data circulation needs multi-subject cooperation. Whereas in the context of the international boundary of data is not clear, the legal framework and rules of data in various countries have not been unified, there are barriers set up for the cooperation of data circulation, which also increases the risk of losing control after data flows out of the territory. What's more, because of the character of portability and reuse, data may face the risk of losing control once it flows out. After the emergence of various privacy protection technologies, although the plaintext sharing of original data can be avoided, it still involves the provision and use of data. Ensuring controllable and traceable data circulation is essential to a secure and efficient data circulation system in Web 3.0 to solve the data disorderly circulation problem.

4 Intelligent and Intrinsic Security Data Circulation System Design

The development of Web 3.0 has brought revolutionary changes in productivity and generative relations, and in the data circulation accordingly. To promote the rational allocation of data resources and return data value to the public, we exploit to design an intelligent and intrinsic security data circulation system based on the infrastructural enablers for Web 3.0, which help any data subject realize an efficient, low-cost, trust and secure data circulation to share data value.

4.1 Participation Subject

Data circulation has the characteristic of multi-subjectivity, which includes data providers, data users, data supervisors, and data evaluators. Since multi-party cooperation is a major feature of the Web 3.0 ecosystem, the role of data subjects should be given full play to promote data circulation, which should consider balancing the interests of various data subjects.

4.1.1 Data Provider

In the Web 3.0 ecosystem, besides the government, enterprises, and organizations, individuals also can be data providers. the data provision usufruct considers not only the enterprises but also the individuals and government. Public data from the government play a role as the main force for data circulation. In this way, it can stimulate innovation and upgrade data circulation.

4.1.2 Data User

According to data classification and user requirements, the provided data including the original data, anonymized data, and processed data, could be provided in offline mode or in real-time dynamic mode to ensure the timeliness of data. Therefore, the data user chooses to process the data with the data provider by the technique like privacy computation method to provide the data security of both data user and data provider in some scenarios.

4.1.3 Data Evaluator

The data evaluator provides a comprehensive evaluation of the entire lifecycle of the data circulation through multiple dimensions, such as data quality and data security to ensure legality and risk controllability. Through evaluation, it can help the data subject to make up the weak points and improve the capabilities, and then promote more efficient and reliable data circulation.

4.1.4 Data Supervisor

The data supervisor aims to protect the interests of all data subjects, and regulate the order of the data circulation market. By supervising the compliance of the data subject and process behavior, the data supervisor controls the risk and coordinates all subjects to jointly deal with risks to ensure the security and compliance of data circulation.

4.2 System Design

Based on the infrastructural enablers for Web 3.0, we design an intelligent and intrinsic security data circulation system. According to the features analysis of data circulation in Web 3.0, the proposed system is designed in a “production - analysis - decentralized interaction – support application - revenue feedback” hierarchy to provide an efficient, low-cost, secure, flexible, and scalable foundation for data circulation, which is shown in Fig. 2.

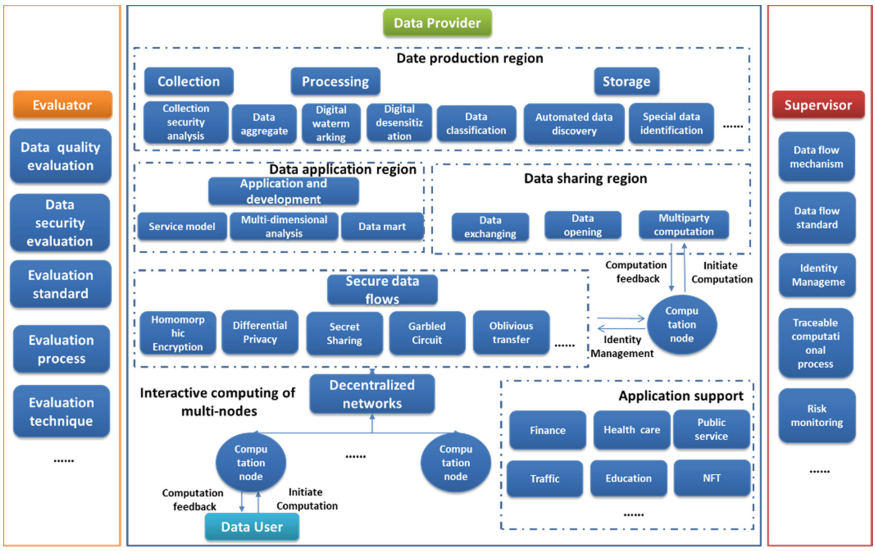

Fig. 2. The framework of the intelligent and intrinsic security data circulation system.

Table 2. Example of the garbled circuit gate table

	$\kappa_{x_1}^0$	$\kappa_{x_1}^1$
$\kappa_{x_2}^0$	$Enc_{\kappa_{x_1}^0}(Enc_{\kappa_{x_2}^0}(\kappa_o^0))$	$Enc_{\kappa_{x_1}^1}(Enc_{\kappa_{x_2}^0}(\kappa_o^0))$
$\kappa_{x_2}^1$	$Enc_{\kappa_{x_1}^0}(Enc_{\kappa_{x_2}^1}(\kappa_o^0))$	$Enc_{\kappa_{x_1}^1}(Enc_{\kappa_{x_2}^1}(\kappa_o^1))$

4.2.1 Hierarchical Management

Based on the data classification result, we set different layer for data processing and circulation. Date production layer process the data collected from various way and chose the proper class based storage strategy. Then Multi-dimensional analysis method are utilized to support data application and development for multiple scenarios in the data application layer. In the data sharing layer, data subject can chose different way to share data with other subject after proper desensitization or encryption processing.

4.2.2 Intelligent Circulation

Based on the decentralized underlying architecture of Web 3.0, data subjects could share and transmit data with other subjects in the system. During this procedure, cryptography, privacy-preserving computation, and distributed ledger technology are utilized to ensure data transmission security. For instance, data subject Alice tries to share data with data subject Bob by privacy-preserving computation based on the Garbled Circuits technique [11].

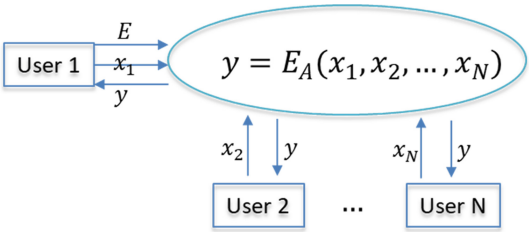

Fig. 3. Multi-subject data sharing model.

The input of Alice and Bob is x_1 and x_2. The processing flow is shown as follows:

1. Alice initiates a data sharing request and generates a multiparty computation $F_A(x_1, x_2)$.
2. Transform F_A to the secure garbled circuit E_A, instead of sending data directly x_1 and x_2 are randomly permutated by 0 and 1.
3. Alice present garbled Boolean circuit gate of E_A, such as shown in Table 2, the wire key of Alice and Bob input and output defined by $\kappa_{x_1}^0$, $\kappa_{x_1}^1$, $\kappa_{x_2}^0$, $\kappa_{x_2}^1$, κ_o^0, κ_o^1, and $Enc_\kappa(.)$ is symmetric key encryption by using key κ.
4. Transmit Garbled circuit E_A and encoded input x_1 and the key to Bob by OT technique thought circulation system.
5. Bob achieve the key according to input x_2 and decryption to obtain the correct garbled gates, which is used to decrypt the next layer of garbled gates.
6. Continue until complete the integrate circuit.
7. Bob share the output $y = E_A(x_1, x_2)$ with Alice.

Through the combination of multiple methods, the system provides an intelligent path for data circulation to meet the personalized need of data subjects. As in Fig. 3, the data interaction model can be extended when there have multiple data subjects.

4.2.3 Application Support

According to the requirements of the various data application scenarios, data scale, and importance of the computing data, the schemes and computing methods are selected to build the efficient and secure data circulation path. Specially, TEE can be utilized when the data value is high and the scale is large. Therefore, the data circulation path in the proposed system can be customized by data subject to support scenarios such as financial risk control and allied medical care.

4.2.4 Identity Management

With more subjects participating in circulation, effective and secure identity management is needed, especially, when facing cross-platform data circulation. Based on the decentralized technology in Web 3.0, the proposed system provide a new mechanism for identity management by combining digital signature, zero-knowledge proof, and trace technologies. As shown in Fig. 4, data subjects participate in the circulation system

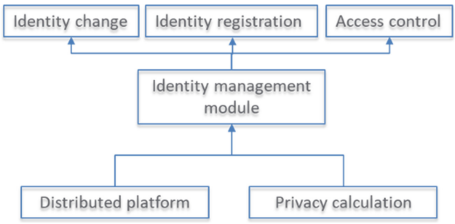

Fig. 4. Identity management data circulation

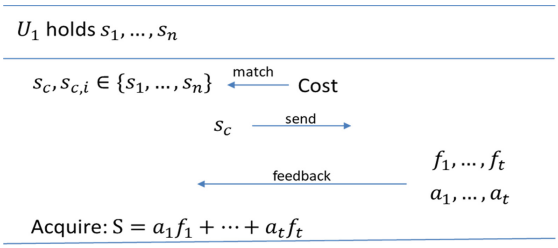

Fig. 5. Security scheme generation model

through registration, and identity management can be refined by a module based on distributed platform and privacy calculation.

In addition, based on real-time identity management tamper-proof storage is carried out in the distributed network for the data processing behavior of the data subject, which provides a credible basis for traceability and authority control to realize the verification, traceability, interpretation, and supervision of the entire data process.

4.2.5 Security Protection

Since the risks such as adversary deception and data eavesdropping in privacy enhancement algorithms, multidimensional security protection is considered in the proposed system. By integrated with artificial intelligence, data leakage prevention, and other new technologies, it provides a dynamic closed-loop intelligent data security capability of the life cycle for data circulation, including monitoring and pre-warning, asset sorting, security protection, compliance management, risk control, partner management, etc. However, overmuch security methods adopted in circulation increase the ability to resist the risk, also reduce the efficiency, and raise the cost. Therefore, the proposed system provides an intelligent way to protect the data in circulation based on the compliance requirements and various scenarios needs. As shown in Fig. 5, data subject U_1 hold the security need $\{s_1, \ldots, s_n\}$, then adjust the security need according to data scale, importance, and cost. Based on the priori knowledge and sample database, system provide the security technique $\{f_1, \ldots, f_t\}$ and weight $\{a_1, \ldots, a_t\}$ from s_c. U_1 can acquire the secure scheme by $S = a_1 f_1 + \cdots + a_t f_t$.

4.2.6 Evaluation and Supervision

To provide more perspective on data circulation to improve the ability and promote development, the system introduces verifiable and measurable data circulation evaluation for various scenarios including data quality and data security. Then to handle the continuous increase of data security threats, the system transforms the traditional peripheral supervision mode to embedded supervision, and the supervision ability is an endogenous part of the data circulation system. The data circulation parameters recorded in the decentralized network are used to supervise malicious behaviors such as fraud and theft in the computing process, which makes the entire data circulation traceable.

5 Conclusion

With the implementation of the Web 3.0 ecosystem, there will be an increasing demand for data circulation and security. In our work, by analyzing the characteristics of data circulation and based on the key categories of infrastructural enablers for Web 3.0, an intelligent and intrinsic security data circulation system is proposed to cope with the new challenges in the Web 3.0 data ecosystem and help to realize the vision of security symbiosis and value sharing. The proposed system enhances the trust in data interaction and attracts more subjects to participate in the new mode of data circulation from six aspects, including hierarchical management, intelligent circulation, application support, identity management, security protection, and evaluation and supervision. Next, the proposed system can be adjusted and applied in a wider range of scenarios to promote the realization of the Web 3.0.

References

1. Yao Qian, Web 3.0: The approaching new generation of Internet, in: China Finance, 2022(06): 14-17.
2. Z. Liu et al, Make Web 3.0 Connected, in: IEEE Transactions on Dependable and Secure Computing, doi: https://doi.org/10.1109/TDSC.2021.3079315.
3. K. Shrishak and H. Shulman, MPC for Securing Internet Infrastructure, in: 2020 50th Annual IEEE-IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S), 2020, pp. 47-48, doi: https://doi.org/10.1109/DSN-S50200.2020.00026.
4. T. Li, A. K. Sahu, A. Talwalkar and V. Smith, Federated Learning: Challenges, Methods, and Future Directions, in: IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, May 2020, doi: https://doi.org/10.1109/MSP.2020.2975749.
5. ZHU Zong-wu and HUANG Ru-wei, Secure Multi-Party Computing Protocol Based on Efficient Fully Homomorphic Encryption, in: Computer Science: 1-13 [2022-07-19].
6. Xiang Zhu, Research and Implementation of Differential Privacy Protection Technology under Federated Learning, in: Nanjing University of Posts and Telecommunication, 2021. DOI: 10.27251/d.
7. LI Y C, ZHOU K J, WANG Z Z, A survey of block chain privacy protection research based on zero-knowledge proof, in: Aerospace Control and Application, 2022, 48 (1): 44-52.
8. XU Qian, ZHANG Qing, YU Bo, YU Wenqing, HE Wei, Middleware and blockchain based interconnecting system of heterogeneous privacy preserving computing platforms, in: Information and Communications Technology and Policy,2021,47(6): 38-49.

9. F. E. Idachaba, 5g networks: Open network architecture and densifcation strategies for beyond 1000x network capacity increase, in: Future Technologies Conference, 2017.
10. TANG Ling-tao, WANG Di, ZHANG Lu-fei and LIU Sheng-yun, Federated Learning Scheme Based on Secure Multi-Party Computation and Differential Privacy, in: Computer Science: 1-14 [2022-07-19].
11. A. C. Yao, Protocols for secure computations, in: 23rd annual symposium on foundations of computer science. IEEE, 1982, pp. 160-164.