Centralized Versus Decentralized Digital Identity Architectures: Simulation Models of Data Exchange

Yoshiaki Fukami,[1] Takumi Shimizu, [2] Teruaki Hayashi, [3]
HirokiSakaji, [4] Hiroyasu Matsushima[5]

Keio University,[1,2] The University of Tokyo,[3, 4] Shiga University,[5]
yofukami@sfc.keio.ac.jp,[1] takumis@sfc.keio.ac.jp,[2]
hayashi@sys.t.u-tokyo.ac.jp, [3] sakaji@sys.t.u-tokyo.ac.jp,[4] hiroyasu-matsushima@biwako.shiga-u.ac.jp[5]

Abstract

In order to utilize big data generated from distributed cloud-based services, a digital ID is required to link between data and its subjects. Decentralized Identifiers (DID) have been developed to manage data from various services with privacy protection. We analyzed two ID architectures, DID and centralized ID (CID), with simulation models to evaluate the efficiency of ID architectures. In a monopoly market where there is no competition between ID providers, there is no difference between DID and CID. However, if there are multiple ID providers without interoperability, service providers have access to more data in the DID architecture compared to CID. However, this result was affected by the design of the model without ID federation technologies. Currently, service providers can receive data from many third-party services with the ID federation standard. Also, the simulation results that DID is very efficient for data distribution should be carefully interpreted by considering the upcoming costs for implementation.

Background

In recent years, consumers have come to have a large number of user accounts linked to more and more cloud-based services. This has led to the accumulation of a wide variety of attribute data in the cloud, increasing the potential for the creation of new services, while at the same time developing a means of sharing data that is fragmented between services in a way that is easy to use and protects the rights of consumers. Service providers can identify consumers with digital IDs provided by third party companies and obtain attribute data stored by other services under consumer authentication.

Most of the data accumulated from multiple services is linked to the ID issued by a specific small number of companies, and such companies also provide functions of authorization. This means that there is some risk that distributed data could be accumulated, analyzed and utilized for unintended use under malicious intent. The risk of privacy infringement is increased by aggregating various attribute data. While the ID federation enhances consumer convenience, it also increases the risk of privacy breaches.

DID is an architecture in which the entity that provides attribute information issues digital IDs in a distributed manner enabled by blockchain technologies. In contrast to DID, an architecture that uses existing ID federation technology is called a Centralized Identifier (CID). With DID, aggregated data can be utilized only with consumer's authentication, and without linking to specific ID providers such as Google and Facebook.

From the service provider's point of view, it is advantageous to be able to obtain and utilize diverse data at low cost, and it will encourage the emergence of innovations in the form of new services. Both architectures, CID and DID, have their advantages and disadvantages, and it is difficult to determine which is better simply. Therefore, we use a simulation approach in order to study many factors in an integrated manner.

In multi-agent simulation, people and objects can be represented as agents, and phenomena resulting from their interactions can be observed. For example, it is applied to fields such as traffic (Bazzan & Klügl, 2009), pedestrian flow (Yamashita et al., 2014), and market transactions (Hirano et al., 2020; Yagi et al., 2020). By confirming the simulation results, it is possible to support decision-making in planning and policy making related to them.

Models

This study employs simulation models to analyze the CID and DID structures and their impacts on data exchange. In the CID model, each user has some data which is managed by ID providers. Service providers have their needs (i.e., which data a service provider needs to create products) and try to obtain the data they need by accessing the IDs users have. Verifiers may or may not get the data depending on an ID that bridges transactions between users and verifiers. For instance, if a verifier asks a user to share the data “a” and the user uses the ID “A” for this transaction, the verifier can get the data “a”. If the user uses the ID “B” in this case, the verifier cannot get the data. In the DID model, there is no ID provider in the transaction. A verifier directly contacts a user and requests the data it needs. Each user decides whether he/she accepts the request from a verifier. These models aim to uncover the efficient data exchange structure considering various parameters such as the number of users and CID providers and the cost of transactions. Figure 1 describes the model structures.

In T. Kido, K. Takadama (Eds.), Proceedings of the AAAI 2022 Spring Symposium “How Fair is Fair? Achieving Wellbeing AI”, Stanford University, Palo Alto, California, USA, March 21–23, 2022.

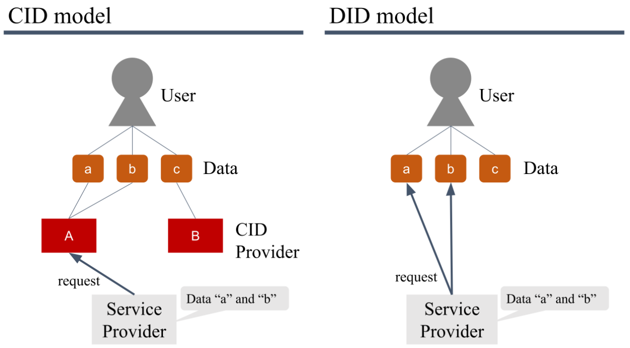

Figure 1: The overview of the models

Results

We evaluate the models based on the number of data that a service provider can access depending on the ID structures. In the CID models, the key parameter is the number of CID providers. If there is one CID provider, a service provider can access all the user data via this particular CID provider. Our simulation assumes 10,000 users in the model, so a service provider can access 10,000 user data in this case. As the number of CID providers increases, user data is dispersed across CID providers and a service provider can obtain only subsets of user data via a CID provider. In the DID models, the key parameter is the attrition rate of service provider’s data request. Since the DID requires users to manage each transaction per data record by themselves unlike the CID which allows CID providers to manage it, a service provider sometimes cannot obtain the data due to this burden of user’s data management. Figure 2 shows the results of our simulation models considering various levels of key parameters. As the graph indicates, the number of data that a service provider can access dramatically decreases as the number of CID providers increases. On the other hand, the number of accessible data in the context of DID stays relatively large even in the case of high attrition rate.

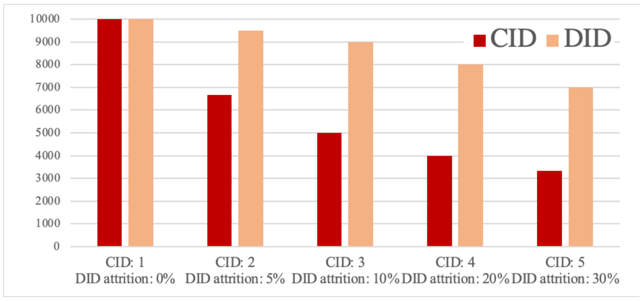

Figure 2: The number of accessible data in CID/DID

Discussion

The result is that service providers have access to more data in the DID architecture compared to CID. However, this result was affected by the design of the model that only introduced the authentication / authorization function of independent third parties without ID federation technologies. Currently, service providers are able to receive data from many third-party services with the ID federation standard such as OpenID connect.

On the other hand, the simulation results show that DID is very positive for data distribution. However, DID has not been diffused yet, and it costs for both data providers and acquirers to implement DID technology. The benefits of DID architecture may be offset or negated by the costs of dissemination, which are not reflected in this model.

Future research needs more fine-grained models which reflect real-world ID operations and practices being developed at standard developing organizations and issues mentioned above such as ID federation and cost structures of ID architectures. This study opens up new research avenues for digital identity structure and data exchange by showing a basic understanding and implications of CID versus DID architectures.

Acknowledgments

This work was supported by JSPS KAKENHI 19K23235, 20H02384, and 20K13599.

References

Bazzan, A.; and Klügl, F. (Eds.). 2009. *Multi-Agent Systems for Traffic and Transportation Engineering*. IGI Global. doi.org/10.4018/978-1-60566-226-8

Hirano, M.; Izumi, K.; Matsushima, H., and Sakaji, H. 2020. Comparing Actual and Simulated HFT Traders’ Behavior for Agent Design. *Journal of Artificial Societies and Social Simulation*, *23*(3). doi.org/10.18564/jasss.4304

Yagi, I.; Masuda, Y.; and Mizuta, T. 2020. Analysis of the Impact of High-Frequency Trading on Artificial Market Liquidity. *IEEE Transactions on Computational Social Systems*, *7*(6): 1324-1334. doi.org/ 10.1109/TCSS.2020.3019352.

Yamashita, T.; Matsushima, H.; and Noda, I. 2014. Exhaustive analysis with a pedestrian simulation environment for assistant of evacuation planning. *Transportation Research Procedia*, *2*: 264–272. doi.org/10.1016/j.trpro.2014.09.047