Framework for Data Trust using Block-Chain Technology and Adaptive Transaction Validation

P. Hari Babu[1], J. Anusha[2], J. Jaya Prakash[3], Ch. Krishna Veni[4], M. Anand[5]
Assistant Professor, Department of Computer Science and Engineering[1]
Under graduate students, Department of Computer Science and Engineering [2,3,4,5]
Raghu Institute of Technology, Dakamarri, Visakhapatnam, A.P. India

Abstract: *Trust is the main barrier preventing widespread data sharing. The lack of transparent infrastructures for implementing data trust prevents many data owners from sharing their data and concerns data users regarding the quality of the shared data. Blockchain technology proposes a distributed and transparent administration by employing multiple parties to maintain consensus on an immutable ledger. This project presents an end-to-end framework for data trust to enhance trustworthy data sharing utilizing blockchain technology. We also suggest an adaptive solution to determine the number of transaction validators based on the computed trust value.*

Keywords: Blockchain, data trust, data sharing, distributed, access control

I. INTRODUCTION

Data sharing has become a big concern regarding privacy and confidential issues, abusing data, and legal and ethical violations. The lack of a transparent and trustworthy framework for data trust hinders many data owners from sharing their data, which could be vital for many research purposes.Data sharing is not merely a big concern for data owners, but also data users are concerned about the trustworthiness and reliability of the provided data at the origin. Hence, trust is a two-way problem for both data owners and data users. Data trust is a fairly new concept that aims to facilitate data sharing by forcing data users to be transparent about the process of sharing and reusing data. Data trust entails legal, ethical, governance and organizational structure as well as technical requirements for enabling data sharing.

Blockchain technology has salient potential to effectively present the essential properties for creating a practical data trust framework by transforming current auditing practices and automatic enforcement of smart contracts logic, without relying on intermediaries to establish trust. Many other studies have investigated blockchain potential for data sharing, establishing trust and access control.. Blockchain can be used as a data trust interface between data controllers and data users. The distributed, secure and reliable nature of the blockchain can reinforce the trustworthiness of the data trust framework.

II. LITERATURE REVIEW

Recently,Shala et alestablished a reward system to encourage IoT network peers with low trust scores to raise it. The motivational system makes use of control loops with a goal trust score. A bundle of incentives, such as discounts for other services, will be provided to service providers with low trust ratings to entice them to deliver a better service in return for the promised advantages. In, authors introduced an incentive-based strategy to motivate medical data owners to share their high-quality (actual and practical) data and receive income, as well as miners who profit by taking part and confirming transactions.

Wang et al. developed a system for an incentive that protects anonymity in order to generate high-quality crowd sensing contributions. Participants are encouraged by the trust mechanism to give their high-quality sensing data in exchange for Bitcoin or Monero. Data miners also make money by ensuring the accuracy of the data.

Dedeoglu et al. provided a trust model to evaluate the accuracy of data collected by IoT network sensor nodes. The credibility and repute of the data source, together with evidence from observations made by other neighboring sensor nodes, make up the model.

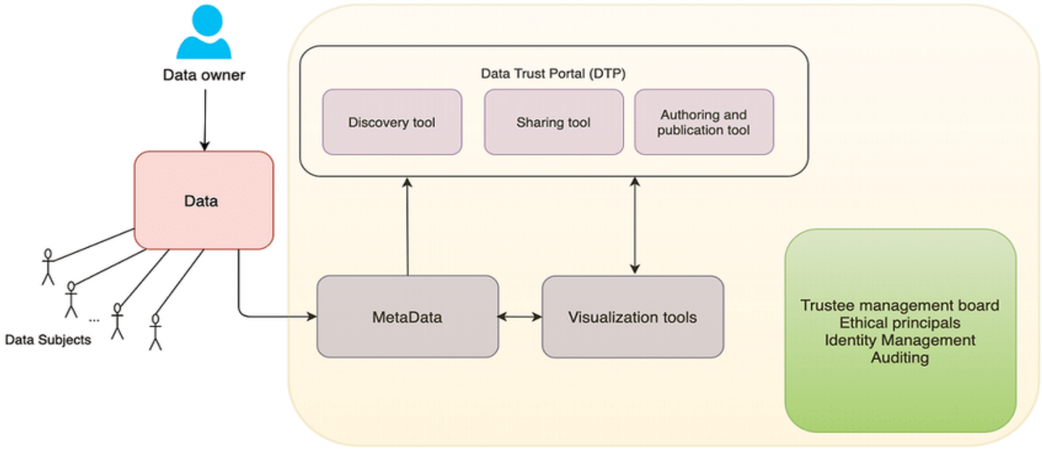
DOI: 10.48175/IJARSCT-9272

III. DATA TRUST CONCEPT

Trust is a multidisciplinary and multifaceted concept that has been defined in various disciplines, such as sociology, economics, psychology, computation, information and computer science, to model different types of relationships. Trust's definition can be challenging since it embraces the facets of ethics, emotions, values, and various disciplines. Fundamentally, Trust is a relationship between trustor and trustee in which the trustor relies on the trustee based on a given criterion. Summarize the trust definition after studying trust across multiple disciplines, as the inclination of the trustor to accept a subjective belief that a trustee will exhibit responsible behavior to maximize the trustor's interest under uncertainty of a particular situation based on the cognitive assessment of previous experience with the trustee.

Typically, digital trust is considered a computational value established from a relationship between trustor and trustee, and measured by trust parameters and evaluated by a defined mechanism. It is essential to understand the interaction between different actors and components and establish mutual trust. O'Hara proposes eight essential properties that underlie data trust architecture [1], (1) discovery, (2) provenance, (3) access controls, (4) access, (5) identity management, (6) auditing of use, (7) accountability, (8) impact. He proposed the Web Observatory as a candidate technology to carry out the required operation of a data trust.

- Discovery refers to the process of discovering the quality and properties of data by data users in the first place.
- Provenance refers to the ability of data users to access the historical record and metadata about the data.
- Access control refers to the ability of data owners to control and manage access permissions toward their data.
- Access refers to the mechanism that provides access for data users.
- Identity management refers to the ability of data owners to identify and authenticate data users.
- Auditing of use refers to providing a transparent history of data usage.
- Accountability refers to achieving accountability by access control and auditing of use.
- Impact refers to assessing the value, usage and misuse of data through recorded information in the data trust.



Data Trust

3.1 Data Trust Portal Architecture (DTP)

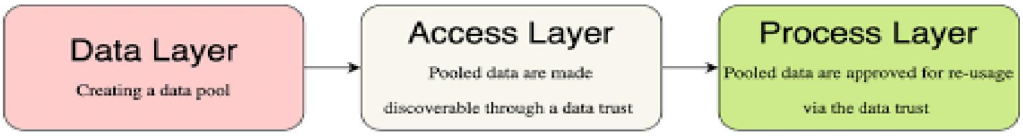
Figure illustrates O'Hara's architecture for data trust called Data Trust Portal (DTP) for sketching out the above features inspired by web observatory infrastructure. The data does not store in DTP, data owners hold the data, and they are responsible for their data protection and the implemented interface method to provide access. DTP is a platform for implementing secure discovery and sharing protocol using metadata about data properties and its provenance.

Stalla–Bourdillon et al. presents a workflow that addresses data protection by design approach to achieve effective data usage, sharing, and reusing. The authors emphasize that such a design requires well-defined data governance roles and processes. They represent data trust through three core layers as represented in figure:

(1) The data layer

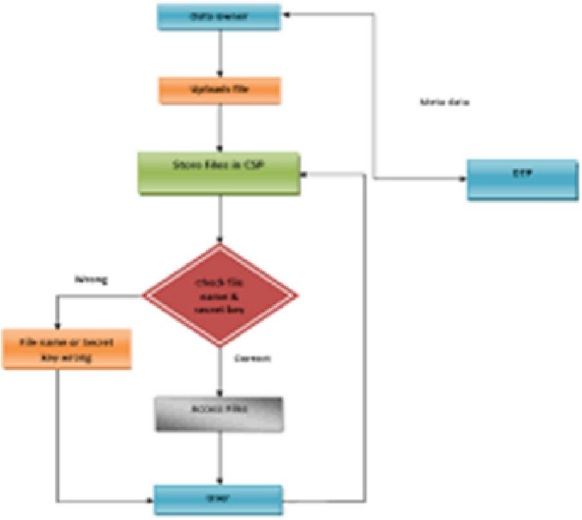
(2) The access layer

(3) The process layer


Data Layer
Creating a data pool
Access Layer
Pooled data are made discoverable through a data trust
Process Layer
Pooled data are approved for re-usage via the data trust

The workflow of data protection by design approach

In the data layer, interested parties sketch out the step of creating data pools by making sure that everyone is aware of the legal requirements for data protection by design, specifying the authorized individuals to decide and act on the pooled data, and preparing data for sharing by applying technical procedures to remove personal identifiers, such as de-identification and anonymization. In the access layer, the data become discoverable for eligible parties by specifying standardized access through centralized or peer to peer technical solutions complemented by monitoring and auditing processes. In the process layer, the pooled data are approved for re-usage via the data trust. This layer is responsible for controlling data usage through standardized risk assessments and ensuring that data are tailored to queries.



Flow Chart

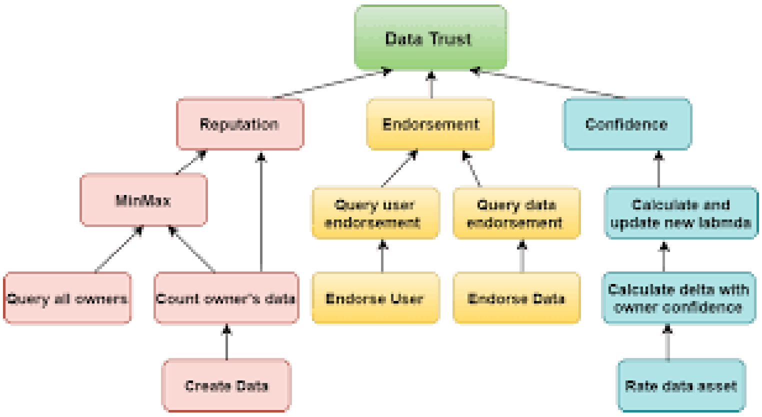
IV. SYSTEM ARCHITECTURE

As we discussed earlier, our proposed system aims to establish a data trust framework beneficial for both data owners and data users. To tackle this goal, we present two main components in our system architecture. 1) a trust model to examine the quality of input data sets and 2) a secure and traceable access control management. Figure presents our data trust framework architecture. We model trust for the input data sets. For any initial data set, our system calculates its trust value through a blockchain based application.This value is used to ensure only trusted data sets are confirmed, and the system only records trusted data assets on the ledger. Section V explains which parameters are involved and how the trust value is calculated. The data sets with lower trust values are considered suspicious, and they are required to be validated by more verifiers.

Once the data sets information is recorded as data assets on the ledger, the data users interested in accessing a data set can prepare a request to access the data set. The data owner will receive the requests directly, and they will decide to give access to their data sets under which terms and conditions. Using blockchain and exploiting smart contracts, all transactions are automatically enforced, and there are no third parties involved. The access control policies could be served through a smart contract, created and stored on the blockchain directly by the resource owner. Moreover, the data owners can transparently query the immutable and permanent storage of blockchain and trace those who had

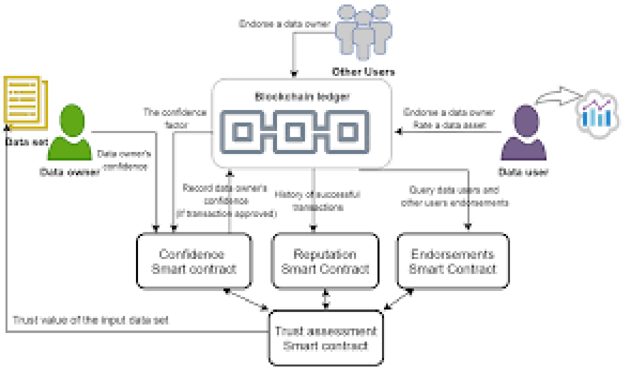
DOI: 10.48175/IJARSCT-9272

access to their data set previously and currently have access to the history of access requests despite the data owners' response type.



V. TRUST MODEL

In this section, we discuss how we calculate the trust value for the input data sets. Later, this value will be available for the data users interested in a particular data set. Moreover, the system adaptively includes this value to define the number of verifiers for confirming the data set. The higher trust value requires fewer verifiers; therefore, the data set will be verified faster.


Endorse a data owner
Other Users
Blockchain ledger
The confidence factor
Endorse a data owner
Rate a data asset
Data set
Data owner
Data owner's confidence
Data user
Record data owner's confidence (if transaction approved)
History of successful transactions
Query data users and other users endorsements
Confidence Smart contract
Reputation Smart Contract
Endorsements Smart Contract
Trust value of the input data set
Trust assessment Smart contract

5.1 Terminology

We call data sets data assets as they are assets that we use in our distributed data trust framework to manage them and control access to them. Every data asset has an owner (data owner) that has full control over the data. Every data asset has a unique identifier (key). Data owners are the one who provides the data asset. The data owner has full control to decide who will have access to the data, for what purpose and the access permission level and conditions. The system provides a transparent history of all granting and revoking access permissions executed by the data owner.

5.2 Reputation

The data owner's reputation is calculated considering the user's previous successful transactions as well as the minimum and the maximum number of successful transactions for all users, using the min-max normalization.

5.3 Endorsement

Data owners can receive two types of endorsements. In the first type, any user in the system who knows the data owner can endorse one; for example, they could have worked together previously. The second type of endorsement is received from the data users who have previously studied a data set provided by the current data owner. The data owner receives an endorsement based on the data user's experience regarding the data set's high quality. The second type of endorsement has a more substantial influence on the data owner's total endorsement score, which can be defined by α factor

DOI: 10.48175/IJARSCT-9272

5.4 Confidence

For any initial data set, the data owner enters a confidence value between 0 to 1 (considenceiϵ (0, 1]) to express its confidence in the provided data set. This value will only be considered in the data asset's total trustworthiness if there is a history of previously entered data sets by the current data owner that was studied by a data user

VI. SYSTEM RQUIREMENTS

6.1 H/W System Configuration

- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor – SVGA

6.2 Software Requirements

- Operating System - Windows XP
- Coding Language - Java/J2EE(JSP,Servlet)
- Front End - J2EE
- Back End - MySQL

VII. ACCESS MANAGEMENT AND SHARING DATA ASSETS

As discussed previously, our end-to-end data trust framework addresses both data owners' and data users' concerns. In the previous section, we explained how our data trust model calculates the trustworthiness of input data sets and how it adjusts data owners' confidence in their provided data sets. This section describes how we can implement a secure and trustable access management system using distributed ledger technology. We demonstrate how to design smart contracts to meet the requirements of the data trust framework.

Blockchain's features, such as transparency, auditability and trust distribution, along with leveraging smart contracts, make it possible to achieve secure and fine-grained access control, thereby promoting the data-trust framework. This section introduces our access control and consent management components for the presented data trust framework

Access control smart contract receives access requests, checks default access permissions. Suppose the user has access to the data set based on the system's previous rules and prior consent between the data owner and data user. In that case, it submits a transaction for recording the access request and the result of access permission.

If there are no default access rules that match the current access request, it sends the data owner's access request. The data owner investigates the access request and decides to accept or reject it based on metadata provided by the data owner. Suppose the response received from the data owner agrees with the access request. In that case, the transaction invokes the consent management smart contract to handle the agreement between the data owner and the data user

VIII. RESULTS

9.1 Discovery

Authenticated data users are able to discover the available data assets, the properties of data sets represented as metadata through the system interface. The information related to data assets' quality is also available for the data users through the trust value calculated by our proposed method.

9.2 Provenance

Once the data owners add their data sets as data assets to the system, they must attach metadata related to the data provenance, such as the data origin, collection 44 time, and collection method. This information can help both transaction verifiers and data users to assess the quality of the data. Moreover, every time a data set is modified, an associated transaction is generated to update the data asset properties on the ledger. It helps to query data provenance and trace data evolution by identifying actual operations that have been performed on the data sets.

9.3 Access Control

Data owners have full control over their data assets. They are the ones who decide on who gets access to their data and by exploiting smart contracts, their access management is enforced automatically. Smart contracts also enable fine grained access checks to verify the authenticity of submitted transactions.

9.4 Access

The data sets that include personal data must be de-identified or anonymized before sharing to ensure that individuals' interests are not compromised by providing access to their information. Besides, Hyperledger Fabric supports private data and private communication, which could be desirable for the data owners who do not want to expose the metadata associated with their data to all system users. They can exploit this feature and share their data assets info with their interested parties. Access to the data provenance can also be limited through customized policies in the smart contracts. For example, data users can send requests to data owners to access reading the data set provenance.

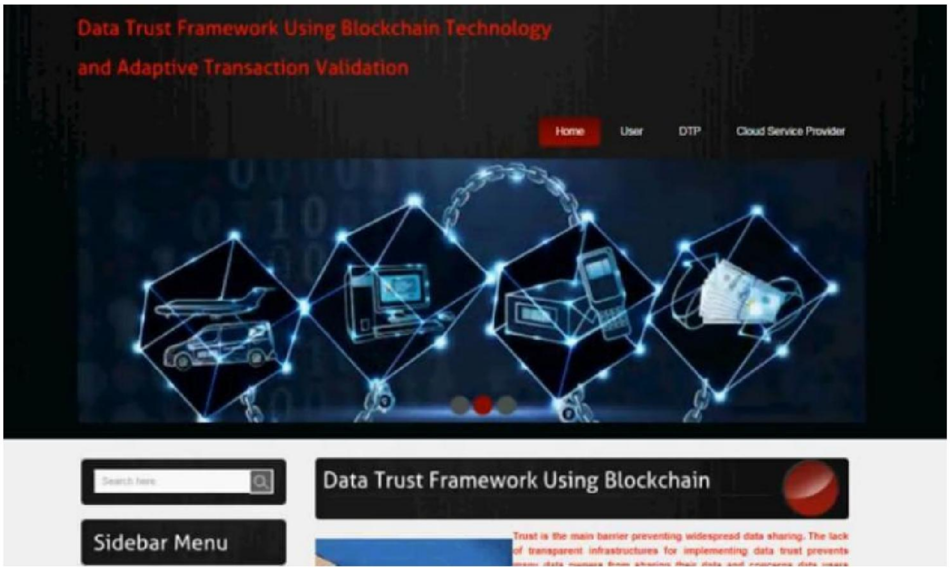
9.5 Identity Management

In Hyperledger Fabric as a permissioned blockchain, a digital identity encapsulated in an X.509 digital certificate must be issued for every actor and user before interacting with the blockchain. This identity is essential to determine the correct permissions over resources and access to information users have in a blockchain network. A digital identity can include additional attributes to specify the person or organization holding that identity. These attributes help data owners to identify those attempting to get access to their data assets

9.6 Auditing

Auditing is one of the primary purposes of introducing blockchain to implement a data trust framework. Blockchain enables us to audit every process and interaction in the system. In the context of data sharing, blockchain provides an immutable audit trail of data modifications, access requests, access grants and revocations. Data owners are able to query the history of transactions regarding access requests and modifications on access permissions on their data assets. Data users are also able to audit data assets origin and history of updates on the data sets. Furthermore, monitoring the immutable log of data transactions to automatically generate audit trails and record any data breach attempts facilitates detecting possible threats

X. SYSTEM IMPLEMENTATION

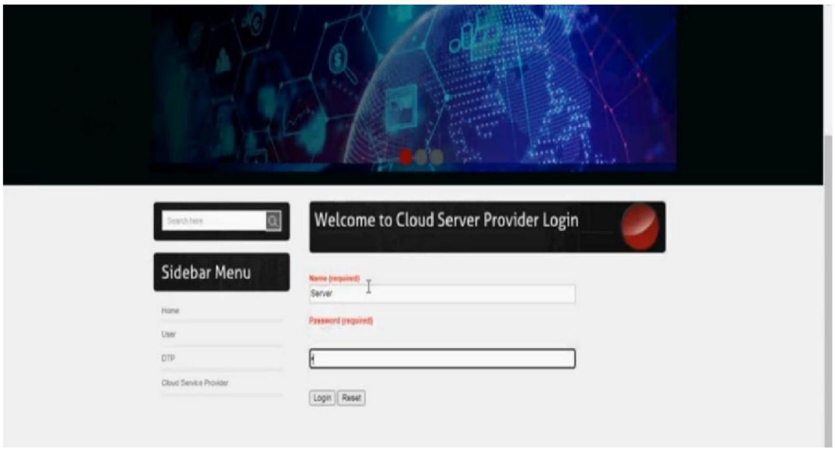


Data Trust Framework Using Blockchain Technology
and Adaptive Transaction Validation

Data Trust Framework Using Blockchain

Trust is the main barrier preventing widespread data sharing. The lack of transparent infrastructures for implementing data trust prevents

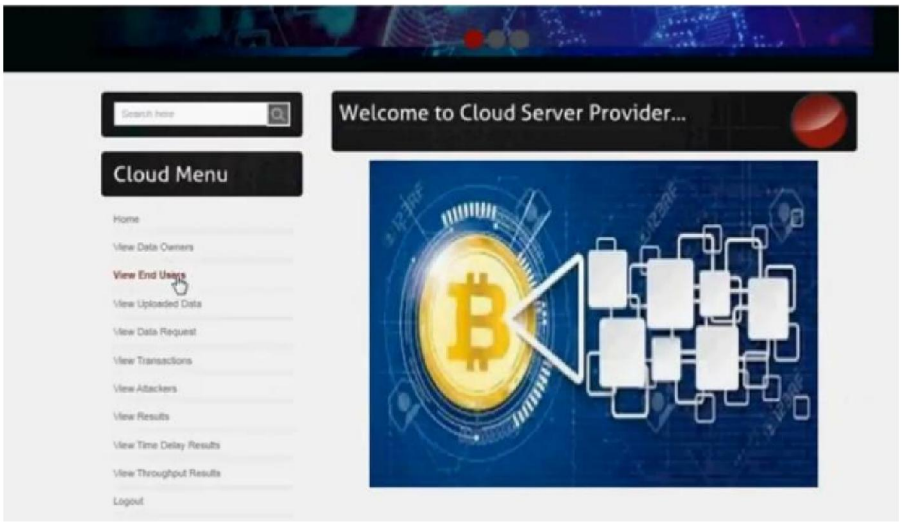
DOI: 10.48175/IJARSCT-9272



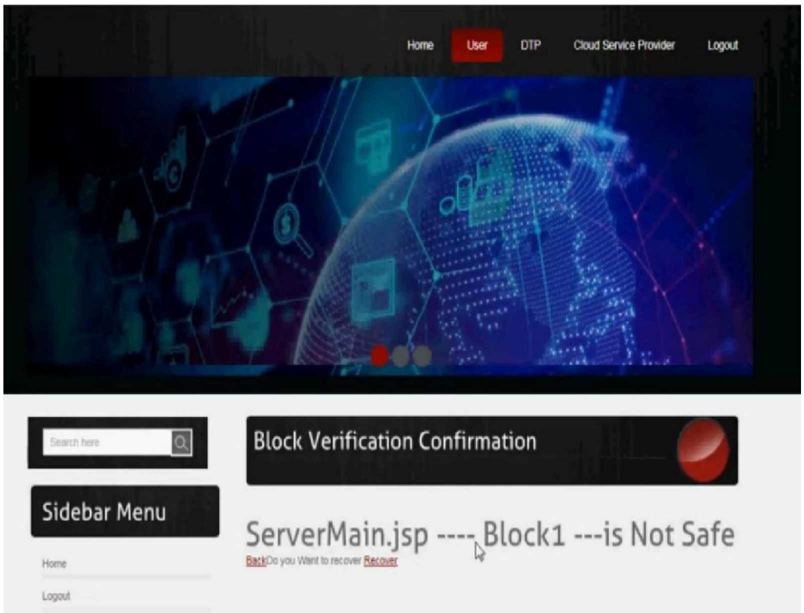
Welcome to Cloud Server Provider Login

Name (required)
Server

Password (required)



Welcome to Cloud Server Provider...



Block Verification Confirmation

ServerMain.jsp ---- Block1 ---is Not Safe

BackDo you Want to recover Recover

XI. CONCLUSION

Current systems are limited in providing a practical and transparent approach for data sharing due to the lack of trust in both parties. In this paper, we introduced an end-to-end data trust framework using permissioned blockchain. Our designed framework assesses the quality of input data using a novel trust model, including the data owner's reputation, endorsements and confidence in provided data. Therefore, the data users ensure that the available data set's quality has been adaptively examined and updated

Data owners also can benefit from secure, transparent, and automatic access management using smart contracts. They have full control over their data assets, and they are the only actors in the system who can regulate access permissions without relying on third parties. By exploiting blockchain's provenance and audibility, data owners can monitor the trace of access regulations and modifications on their data assets. Moreover, valuable logs can be extracted from the ledger to present a transparent view of the system, identify suspicious requests, and detect protocol breaches leading to discovering possible threats. Evaluation results indicate the system's effectiveness in handling a large number of transactions for writing, updating, and querying trust parameters value

REFERENCES

[1]. K. O'hara, ``Data trusts: Ethics, architecture and governance for trustworthy data stewardship," Univ. Southampton, Southampton, U.K., Tech. Rep., 2019.

[2]. A. Alsaad, K. O'Hara, and L. Carr, ``Institutional repositories as a data trust infrastructure," in Proc. Companion Publication 10th ACMConf.Web Sci., Jun. 2019, pp. 1_4.

[3]. S. Rouhani and R. Deters, ``Security, performance, and applications of smart contracts: A systematic survey," IEEE Access, vol. 7, pp. 50759_50779, 2019.

[4]. J.-H. Cho, K. Chan, and S. Adali, ``A survey on trust modeling," ACM Comput. Surv., vol. 48, no. 2, pp. 1_40, Nov. 2015. SPECIALUSIS UGDYMAS / SPECIAL EDUCATION 2022 2 (43) 1061.

[5]. Z. Yan and S. Holtmanns, ``Trust modeling and management: From social trust to digital trust," in Computer Security, Privacy, and Politics: Current Issues, Challenges, and Solutions. Hershey, PA, USA: IGI Global, 2008, pp. 290_323.

[6]. S. Stalla-Bourdillon, G. Thuermer, J. Walker, L. Carmichael, and E. Simperl, ``Data protection by design: Building the foundations of trustworthy data sharing," Data Policy, vol. 2, pp. 1_10, Jan. 202

DOI: 10.48175/IJARSCT-9272