International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

A Data Storing and Sharing Solution with Guaranteed Reliability

Punugupati Sai Kumar[1], Dandu Srivamsi[2], Dr. M. D. Anto Praveena[3]
UG Student[1&2] ,Professor[3]

Department of Computer Science, Sathyabama Institute of Science & Technology, Tamilnadu-India

Article Info
Volume 9, Issue 5
Page Number : 291-300

Publication Issue
September-October-2022

Article History
Accepted : 01 Oct 2022
Published : 09 Oct 2022

ABSTRACT

Digital data certified by a reputable organization are valuable digital data that can be stored or shared on the internet. However, the problems are: (1) How to ensure the anonymity of organizations on issued certificates? (2) How to ensure that valuable digital data are securely stored in the system? and (3)How could people verify the reliability of shared data while still ensuring the confidentiality of its content, and how to ensure that the data sharing process is safe, transparent, and fair? Therefore, we propose data producing, data storing, and data sharing schemas. In the data producing schema, we deploy a group signature scheme for a group of reputable organizations that provide the same type of service, an organization in the group generates a valuable digital data from raw data sent from a data owner and then issues a certificate on the ciphertext of this digital data. In the data storing schema, the data owner uploads his/her data to the public Inter-Planetary File System network and then stores the access address of the stored data and the corresponding certificate on the blockchain ledger. In the data sharing schema, everyone on the system could verify the reliability of shared data before sending a data sharing request to the data owner. The data sharing process is performed via a smart contract, and involved parties have to escrow to encourage honesty. The schemas of data storing and sharing guarantee the security properties including confidentiality, integrity, privacy, non-repudiation, and anonymity.

Keywords: Blockchain, IPFS, data storing, data sharing.

I. INTRODUCTION

There has been exponential data growth in the world, and trusted data are considered one of the most valuable assets of individuals and organizations. The amount of data created and stored globally are predicted to create about 175 zeta bytes by 2025. It is also estimated that by 2025 the global consumers interacting with data everyday will reach 5 billion [1]. Consequently, the demand for valuable data storing and sharing is tremendous, which also poses challenges related to data security in the processes of data storing and sharing. Currently, there are two main architectures used for data storing and sharing, centralized and decentralized architectures

For the centralized architecture, organizations can store data on their data center system. However, these systems have high operating costs and are limited in scalability. Using cloud storage services can reduce costs and can be flexible in system expansion, and more suitable for IoT systems. The combination of IoT and cloud storage services is a matter of studies. To protect the security and privacy of data storing and sharing, encryption algorithms and access control models are proposed, Murat Kantarcioglu et al. proposed SECUREDL for protecting the sensitive data stored in databases. However, the centralized architecture has two limitations including: (1) data security, stored data could be accessed, modified, or removed illegally by system administrators or attackers who compromised the system; (2) availability, when the centralized systems are crashed due to system overload, denial-of-service or distributed denial-of-service (DoS/DDoS) attacks, or system errors, the services are not available for users.

For the decentralized architecture, most solutions use blockchain (BC) technology as the main component in the systems because of its properties such as anonymity, transparency, decentralization, and auditability. However, current solutions do not provide features for verifying the accuracy and the reliability of the shared data on the BC network. Specifically, data verified and certified by a reputable organization (RO) are considered as meaningful data (MD). For instance, in the medical field, a diagnostic result of an electronic medical record is published by a reputable medical organization with highly skilled doctors, which is MD. In the education field, a lecture that is assessed and certified by a professional board of a reputable university is MD. MD needs to be securely stored on the system, besides a data owner (DO) can completely share or commercialize his/her MD to other people or organizations on the network. Data sharing methods must ensure that requesters can verify the reliability and accuracy of shared data before deciding to perform a data-sharing contract. The accuracy and the reliability of the shared data on the BC network. Specifically, data verified and certified by a reputable organization (RO) are considered as meaningful data (MD). For instance, in the medical field, a diagnostic result of an electronic medical record is published by a reputable medical organization with highly skilled doctors, which is MD. In the education field, a lecture that is assessed and certified by a professional board of a reputable university is MD. MD needs to be securely stored on the system, besides a data owner (DO) can completely share or commercialize his/her MD to other people or organizations on the network. Data sharing methods must ensure that requesters can verify the reliability and accuracy of shared data before deciding to perform a data-sharing contract

With the traditional data sharing method, the integrity of shared data is based on trust between the two partners participating in the exchange process. For example, doctors/hospitals absolutely believe that medical records received from their patients are integrity. In some cases, RO needs to ensure anonymity in MD generated by themselves. And the privacy of DO also needs to be protected as they don't want anyone to know which RO's service they used. In addition, the identities of those involved in the sharing process also need to be anonymous; and shared data need to be verified the reliability while still ensuring the privacy of its content.

Data storing and sharing for certified digital data are very necessary, which requires data storage and sharing solutions that need to meet all of the following requirements:

For data storing: The anonymity of certificate authorities and the privacy of DO on stored data must be protected; stored data in the system must be guaranteed confidentiality and integrity.

For data sharing: Everyone on the system can verify the reliability of shared data before submitting a sharing request to DO. Note that everyone can only verify the reliability of the shared data but cannot read its contents. The data sharing process is done directly between DO and DU without depending on any intermediaries. The system serving data storage and sharing must ensure availability, integrity, and scalability. However, current solutions do not meet all of the above requirements. In this paper, we propose data producing, data storing, and data sharing schemes. We consider RO as a data provider (DP), and DPs providing the same type of service join in a group. In the data producing scheme, a group manager sets up a group of DPs that provide the same type of service. A raw data of DO is produced into MD by a particular DP in the group. Then, DP encrypts MD using a symmetric algorithm along with a secret key. Later, DP generates a certificate on the MD ciphertext (denoted by EMD). Finally, EMD, the certificate, and DP's information will be sent to DO through a secure channel. In the data storing scheme, DO stores EMD on Inter-Planetary File System (IPFS), the access address of EMD on IPFS and related information are stored in a transaction on the blockchain system.

However, current solutions do not meet all of the above requirements. In this paper, we propose data producing, data storing, and data sharing schemes. We consider RO as a data provider (DP), and DPs providing the same type of service join in a group. In the data producing scheme, a group manager sets up a group of DPs that provide the same type of service. A raw data of DO is produced into MD by a particular DP in the group. Then, DP encrypts MD using a symmetric algorithm along with a secret key. Later, DP generates a certificate on the MD ciphertext (denoted by EMD). Finally, EMD, the certificate, and DP's information will be sent to DO through a secure channel. In the data storing scheme, DO stores EMD on Inter-Planetary File System (IPFS), the access address of EMD on IPFS and related information are stored in a transaction on the blockchain system.

II. RELATED WORK

Data center network virtualization: A survey: It present a survey of the current state of-the-art in data center networks virtualization, and provide a detailed comparison of the surveyed proposals. We discuss the key research challenges for future research and point out some potential directions for tackling the problems related to data center design.

An IoT-oriented data storage framework in cloud computing platform: A data storage framework integrates both structured and unstructured data in addition to efficiently storing large amounts of IoT data. In order to store and handle the many forms of data gathered by sensors and RFID readers, this data storage framework is able to merge and expand several databases and Hadoop. Additionally, certain parts are created to enhance Hadoop and provide a distributed file repository that can effectively process enormous unstructured file volumes.

Cloud databases for Internet-of-Things data: The proposed solution includes strategies for expressing common IoT data in the form of key-value pairs, as well as a data pre-processing and sharing mechanism to build a Web of Things connecting HTTP-enabled smart devices such as sensors and actuators with virtual "things" such as services, social networks, and APIs. The platform adopts Mongo DB as database server, provides models and interfaces that help to abstract and adopt different types of data and devices. This addressed the challenges related to data that are dynamic, various, massive, and spatial-temporal (i.e., each sample corresponds to a specific time and location). To provide a uniform storage mechanism for heterogeneous sensor data, the system combined the use of the relational model and the key-value model, and was implemented with a PostgreSQL database. Its multi-layer architecture was claimed to reduce the amount of data to be processed at the cloud management layer. Besides, the work also provided several experiments that showed a promising performance when storing and querying a huge volume of data.

Survey of real-time processing technologies of IoT data streams: The survey on emerging technologies toward the real-time utilization of IoT data streams in terms of networking, processing, and content curation and clarify the open issues. Then we propose a new framework for IoT data streams called the Information Flow of Things (IFoT) that processes, analyzes, and curates massive IoT streams in real-time based on distributed processing among IoT devices.

Uploading and replicating Internet of Things (IoT) data on distributed cloud storage: We model our problem thoroughly based on various parameters such as effective bandwidth of the IoT network, available number and size of data items at each mini-Cloud, and we present our problem as a collection of various sub-problems based on subsets of these parameters. The solution assumes the existence of multiple distributed cloud data centres, called mini-Clouds, among which data can be replicated. We prove that the exact solution to the problem is intractable, and we present a number of heuristic strategies to solve it. Our results show that the performance of any heuristic is bounded by the read and write latency of mini-Clouds and the best we can do is often 12 times the worst we can do for a given number of data items to be uploaded and replicated from the gateways to the mini-Clouds in test setup.

IoT data compression and optimization techniques in cloud storage: Current prospects and future directions: This article presents a detailed survey on different data compression and storage optimization techniques in the cloud, their implications, and discussion over future directions. The development of the smart city or smart home systems lies in the development of the Internet of Things (IoT). With the increasing number of IoT devices, the tremendous volume of data is being generated every single day. Therefore, it is necessary to optimize the system's performance by managing, compressing and mining IoT data for smart decision support systems. In this article, the authors surveyed recent approaches with up-to-date outcomes and findings related to the management, mining, compression, and optimization of IoT data. The authors then discuss the scopes and limitations of present works and finally, this article presents the future perspectives of IoT data management on basis of cloud, fog, and mobile edge computing.

III. METHODOLOGY

Proposed system:

Our proposed scheme for data production, data storage, and information sharing We use a group authentication protocol in the data generating structure for a team of reputable companies that provide the identical sort of service. One of the organizations in the group produces valuable online information from raw data sent by a data owner and then concerns a certificate based on the encrypted message of this digital information.

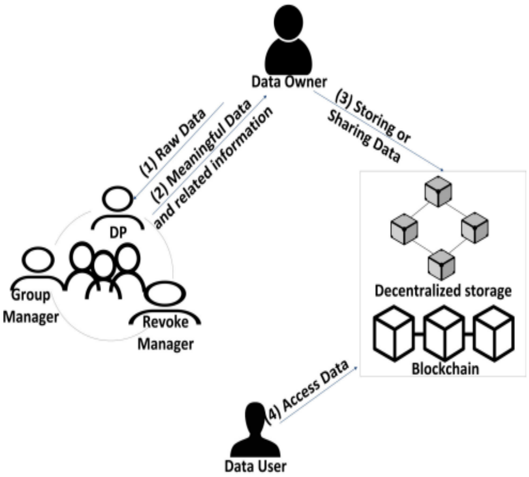


Figure 1: Block diagram

IV. IMPLEMENTATION

The project has implemented by using below listed algorithms.

CLOUD:

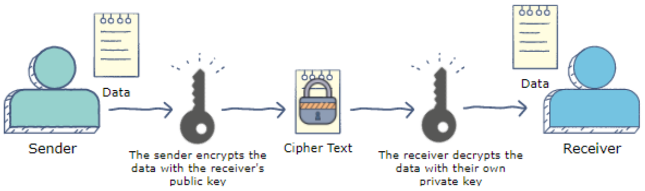
Cloud includes three basic services:

- Infrastructure as a Service (Iaas),
- Platform as a Service (PaaS), and
- Software as a Service (Saas).

Software-as-a-service (SaaS:) entails giving clients access to a software licence. Usually, licences are made available on-demand or on a pay-as-you-go arrangement. This kind of setup is available in Microsoft Office 365.

Infrastructure-as-a-service (IaaS): is a technique for offering anything over IP-based connection as part of an on-demand service, from operating systems to servers and storage. Clients can obtain software and servers through an on-demand, outsourced service rather of having to buy them outright. Microsoft Azure and IBM Cloud are two common IaaS examples.

Platform-as-a-service (PaaS): is said to be the third and most complicated layer of cloud computing. PaaS and SaaS are quite similar, with the main distinction being that PaaS is a platform for developing software that is supplied via the Internet rather than offering software as a service. Platforms like Salesforce.com and Heroku are part of this strategy.


Data
Sender
The sender encrypts the data with the receiver's public key
Cipher Text
The receiver decrypts the data with their own private key
Data
Receiver

DATA ENCRYPTION:

Data is converted into another form, or code, via data encryption so that only those with a secret key (officially referred to as a decryption key) or password may decipher it. Unencrypted data is referred to as plaintext, whereas encrypted data is frequently referred to as ciphertext. At the moment, corporations employ encryption as one of the most common and successful data security techniques. Asymmetric encryption, commonly referred to as public-key encryption, and symmetric encryption are the two primary methods of data encryption.

PURPOSE:

Data encryption is used to safeguard the secrecy of digital data while it is being stored on computer systems and transported over the internet or other computer networks. Modern encryption methods, which are essential to the security of IT systems and communications, have superseded the obsolete data encryption standard (DES).

These algorithms provide crucial security goals like authentication, integrity, and non-repudiation while also ensuring secrecy. A message's origin may be confirmed by authentication, and its integrity can be demonstrated by showing that its contents haven't changed since it was transmitted. Furthermore, non-repudiation guarantees that the sender of a communication cannot retract their actions.

DATA DECRYPTION

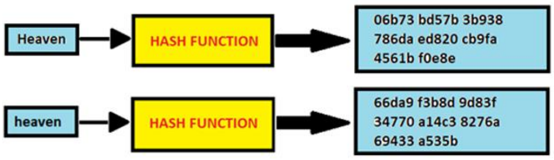
Decryption is the process of restoring the unencrypted version of data after encryption has rendered it unreadable. The system extracts and transforms the jumbled data into sentences and graphics that can be easily understood by both the reader and the system during decryption. The process of decryption can be carried either manually or automatically. A set of keys or a password might also be used to carry it out. Privacy is one of the main justifications for putting in place an encryption-decryption system. Information that is transmitted via the World Wide Web is open to review and access by unauthorized people or organizations. Data is therefore encrypted to decrease data loss and theft. Emails, text files, photos, user data, and directories are a some of the often encrypted goods. A prompt or window asking for a password to access encrypted data is presented to the person in responsibility of decryption.

BLOCK CHAIN

Blockchain is a system of recording information in a way that makes it difficult or impossible to change, hack, or cheat the system. A blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain. Each block in the chain contains a number of transactions, and every time a new transaction occurs on the blockchain, a record of that transaction is added to every participant's ledger. The decentralised database managed by multiple participants is known as Distributed Ledger Technology (DLT).Blockchain is a type of DLT in which transactions are recorded with an immutable cryptographic signature called a hash. This means if one block in one chain was changed, it would be immediately apparent it had been tampered with. If hackers wanted to corrupt a blockchain system, they would have to change every block in the chain, across all of the distributed versions of the chain.

SHA Algorithm:

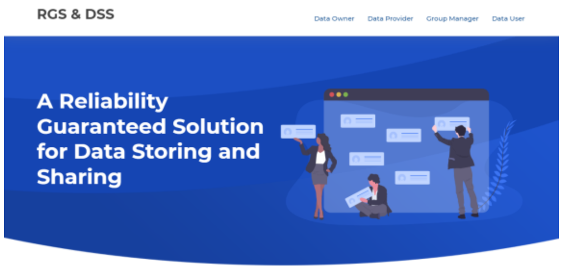
Secure Hashing Algorithm, or SHA. Data and certificates are hashed with SHA, a modified version of MD5. By utilizing bitwise operations, modular additions, and compression functions, a hashing algorithm reduces the input data into a smaller form that is impossible to comprehend. Can hashing be cracked or decrypted, you may wonder? The main distinction between hashing and encryption is that hashing is one-way; once data has been hashed, the resultant hash digest cannot be decrypted unless a brute force assault is applied. To see how the SHA algorithm functions, see the graphic below. SHA is designed to provide a different hash even if only one character in the message changes. For instance, combining two messages that are similar but distinct, such as Heaven and Heaven Is Different The only difference between a capital and tiny letter, though, is size.



V. Results and Discussion

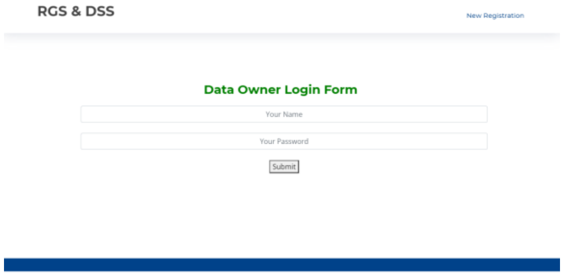
The following screenshots are depicted the flow and working process of project.

Home page: This is the home page of a reliability guaranteed solution for data storing and sharing.



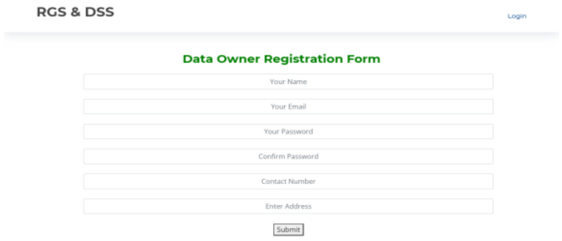
A Reliability Guaranteed Solution for Data Storing and Sharing

Data Owner Login: Data owner can login with valid credentials.



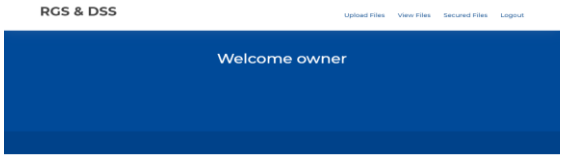
Data Owner Login Form

Data Owner Registration Page: Data owner can register with required details.



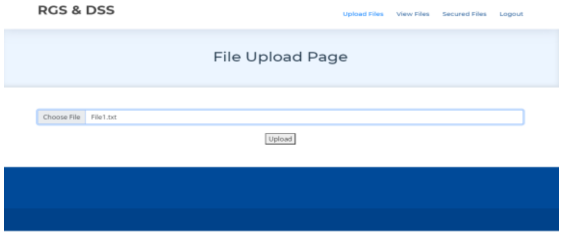
Data Owner Registration Form

Owner Home Page: After successful login data owner can view the home page.



Welcome owner

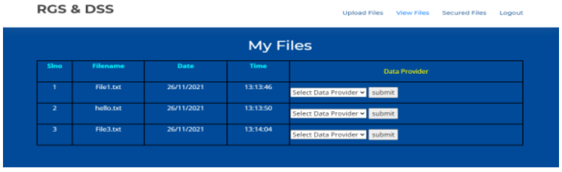
Upload Page: Data can be uploaded here.



File Upload Page

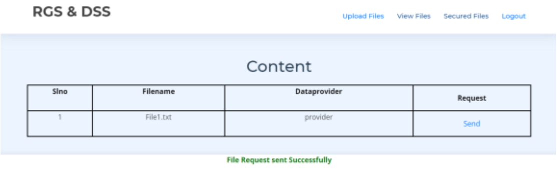
File1.txt

View Files Page: This page displays all the uploaded files.



Slno	Filename	Date	Time	Data Provider
1	File1.txt	26/11/2021	13:13:46	Select Data Provider submit
2	hello.txt	26/11/2021	13:13:50	Select Data Provider submit
3	File3.txt	26/11/2021	13:14:04	Select Data Provider submit

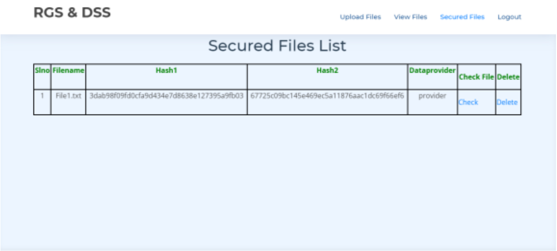
Send Request: This page having requests made by users.



Slno	Filename	Dataprovider	Request
1	File1.txt	provider	Send

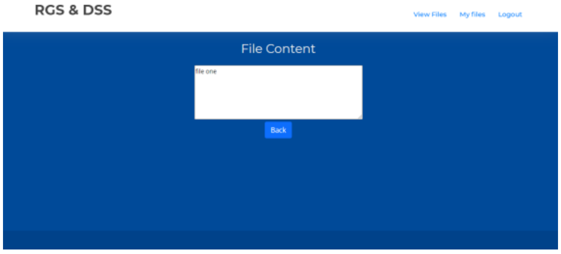
File Request sent Successfully

Data owner secured files: Here we can see the secured files of data owner.



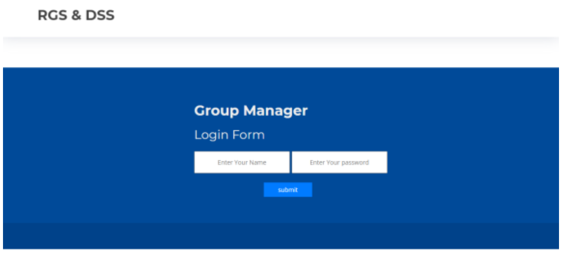
Slno	Filename	Hash1	Hash2	Dataprovider	Check File	Delete
1	File1.txt	3dab98f09fd0cfa9d434e7d8638e127395a9fb03	67725c09bc145e469ec5a11876aac1dc69f66ef6	provider	Check	Delete

File Content: Here we can see the content of selected file.

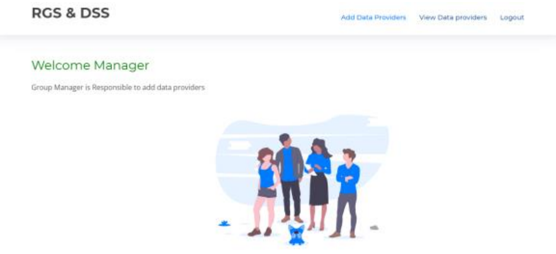


file one

Manager Login Page: Manager can login with valid credentials.

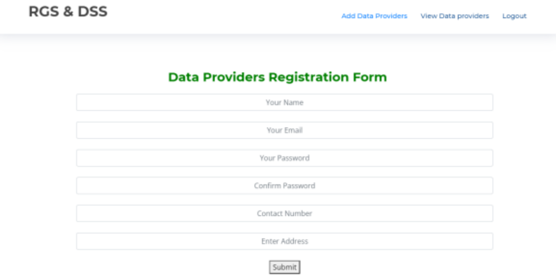


Manager Home Page: Manager can view the home page after successful login.

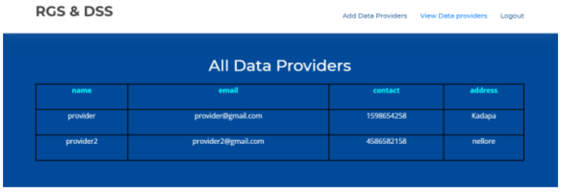


Group Manager is Responsible to add data providers

Add Providers page: Data providers will be added by registering here.

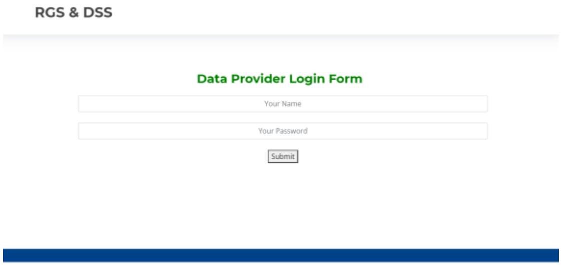


Data Providers list: This page displays all the registered data providers.

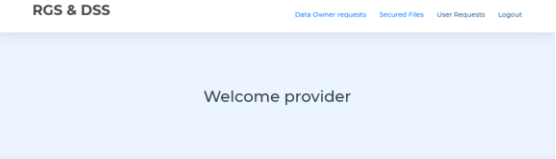


name	email	contact	address
provider	provider@gmail.com	1598654258	Kadapa
provider2	provider2@gmail.com	4586582158	nellore

Providers Login Page: Data providers can login with valid credentials.

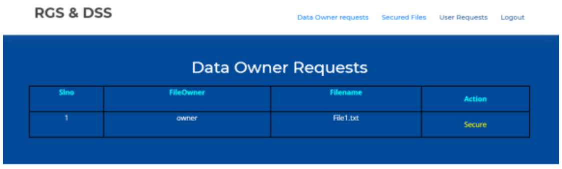


Data providers Home Page: After login data providers can view the home page.



Welcome provider

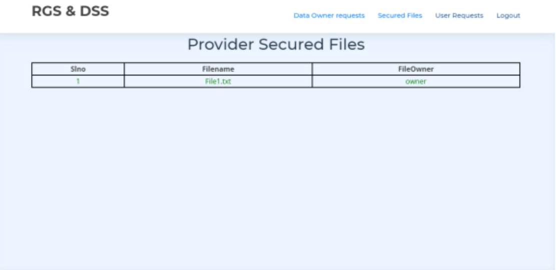
Data Owners Request: Data owner can request for secure file.



Data Owner Requests

Slno	FileOwner	Filename	Action
1	owner	File1.txt	Secure

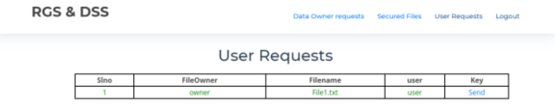
View Secured Files: This page displays the secured file.



Provider Secured Files

Slno	Filename	FileOwner
1	File1.txt	owner

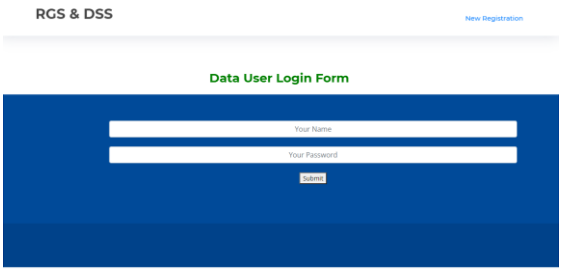
User Requests: User can send the request.



User Requests

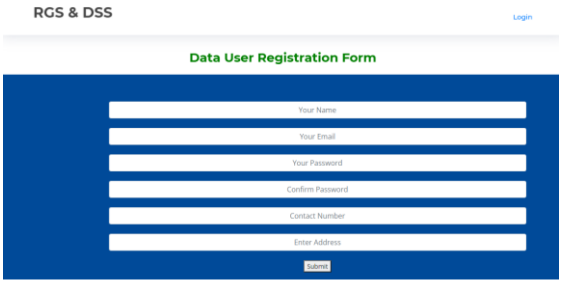
Slno	FileOwner	Filename	user	Key
1	owner	File1.txt	user	Send

Data User Login Page: User can login with valid credentials.



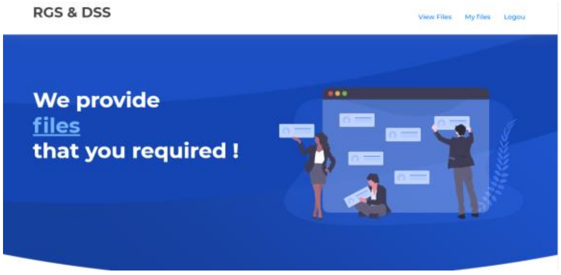
Data User Login Form

Data User Registration: User can register with required details.

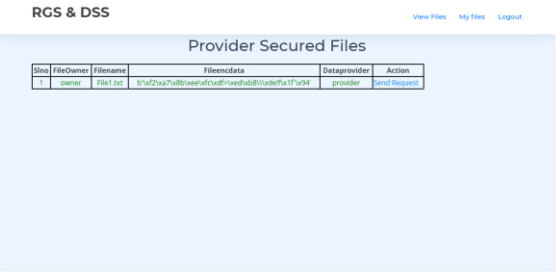


Data User Registration Form

Data User home page: Data user can view the home page after login.



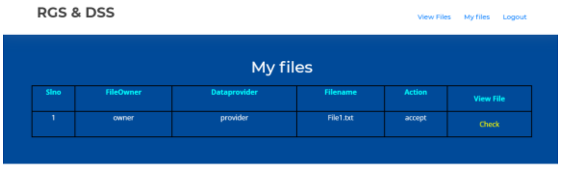
Data User Files: This page contains the files of data user.



Provider Secured Files

Slno	FileOwner	Filename	Fileencdata	Dataprovider	Action
1	owner	File1.txt	b'\xf2\xa7\x8b\xee\xfc\xdf>\xed\xb8\\\xde!f\x1f'\x94'	provider	Send Request

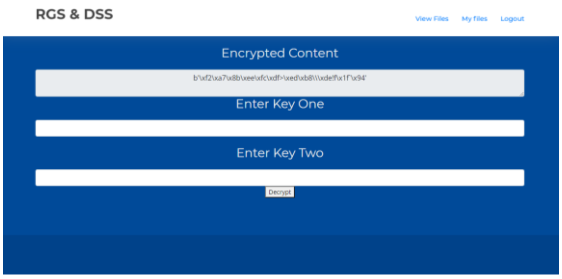
Data User Files: This page contains the content of selected user file.



My files

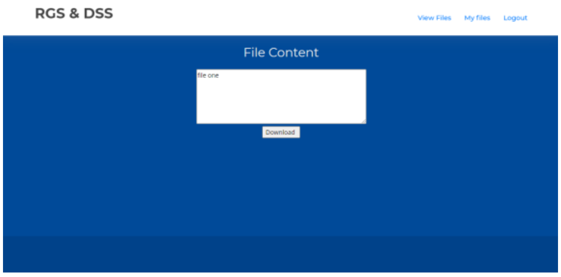
Slno	FileOwner	Dataprovider	Filename	Action	View File
1	owner	provider	File1.txt	accept	Check

Data Decrypt Page: This is the page where we can see the decrypted file.



b'\xf2\xa7\x8b\xee\xfc\xdf>\xed\xb8\\\xde!f\x1f'\x94'

Data User Decrypted Page: Data user can view the decrypted file.



file one

VI. CONCLUSION

In this paper, we propose three schemes: data producing, data storing, and data sharing. In the data producing scheme, we consider RO as DP, a group manager sets up a group of DPs providing the same type of services. DP can generate MD from RD sent from DO, and then issues a certificate on EMD. In the data storing scheme, we provide not only the confidentiality and integrity of the stored data but also the anonymity of DP and the privacy of DO which have not been fulfilled in the existing solutions. In the data sharing scheme, everyone on the system can verify the reliability of shared data before submitting a sharing request to DO. Note that everyone can only verify the reliability of the shared data but cannot read its contents. This property could not be fulfilled by existing solutions. In addition, the data sharing process is done directly between DO and DU without depending on any intermediaries. The results of the security analysis show that the proposed schemes meet the security properties including confidentiality, integrity, privacy, non-repudiation, and anonymity.

VII. REFERENCES

[1]. D. Reinsel, J. Gantz, and J. Rydning, "The digitization of the world from edge to core," IDC White Paper, Nov. 2018.

[2]. M. F. Bari, R. Boutaba, R. Esteves, L. Z. Granville, M. Podlesny, M. G. Rabbani, Q. Zhang, and M. F. Zhani, "Data center network virtualization: A survey," EEE Commun. Surveys Tuts., vol. 15, no. 2, pp. 909–928, 2nd Quart., 2013.

[3]. L. Jiang, L. D. Xu, H. Cai, Z. Jiang, F. Bu, and B. Xu, "An IoT-oriented data storage framework in cloud computing platform," IEEE Trans. Ind. Informat., vol. 10, no. 2, pp. 1443–1451, May 2014.

[4]. T. A. Phan, J. K. Nurminen, and M. Di Francesco, "Cloud databases for Internet-of-Things data," in Proc. IEEE Int. Conf. Internet Things (iThings), IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom), Sep. 2014, pp. 117–124.

[5]. K. Yasumoto, H. Yamaguchi, and H. Shigeno, "Survey of real-time processing technologies of IoT data streams," J. Inf. Process., vol. 24, no. 2, pp. 195–202, 2016.

[6]. A. Kumar, N. C. Narendra, and U. Bellur, "Uploading and replicating Internet of Things (IoT) data on distributed cloud storage," in Proc. IEEE 9th Int. Conf. Cloud Comput. (CLOUD), Jun. 2016, pp. 670–677.

[7]. K. Hossain, M. Rahman, and S. Roy, "IoT data compression and optimization techniques in cloud storage: Current prospects and future directions," Int. J. Cloud Appl. Comput., vol. 9, no. 2, pp. 43–59, Apr. 2019.

[8]. J. D. Bokefode, A. S. Bhise, P. A. Satarkar, and D. G. Modani, "Developing a secure cloud storage system for storing IoT data by applying role based encryption," Procedia Comput. Sci., vol. 89, no. 1, pp. 43–50, 2016.

[9]. W. Wang, P. Xu, and L. T. Yang, “Secure data collection, storage and access in cloud-assisted IoT,” IEEE Cloud Comput., vol. 5, no. 4, pp. 77–88, Jul. 2018.

[10]. M. Rashid, S. A. Parah, A. R. Wani, and S. K. Gupta, “Securing Ehealth IoT data on cloud systems using novel extended role based access control model,” in Internet Things (IoT). Cham, Switzerland: Springer, 2020, pp. 473–489.

[11]. R. Arora, A. Parashar, and C. C. I. Transforming, “Secure user data in cloud computing using encryption algorithms,” Int. J. Eng. Res. Appl., vol. 3, no. 4, pp. 1922–1926, 2013.

[12]. M. Kantarcioglu and F. Shaon, “Securing big data in the age of AI,” in Proc. 1st IEEE Int. Conf. Trust, Privacy Secur. Intell. Syst. Appl. (TPSISA), Dec. 2019, pp. 218–220.

[13]. C. Zhang, J. Sun, X. Zhu, and Y. Fang, “Privacy and security for online social networks: Challenges and opportunities,” IEEE Netw., vol. 24, no. 4, pp. 13–18, Jul./Aug. 2010.

[14]. A. M. Antonopoulos, Mastering Bitcoin: Unlocking Digital Cryptocurrencies. Sebastopol, CA, USA: O’Reilly Media, 2014.

[15]. Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, “Blockchain challenges and opportunities: A survey,” Int. J. Web Grid Services, vol. 14, no. 4, pp. 352–375, 2018.

[16]. M. Conti, E. S. Kumar, C. Lal, and S. Ruj, “A survey on security and privacy issues of bitcoin,” IEEE Commun. Surveys Tuts., vol. 20, no. 4, pp. 3416–3452, 4th Quart., 2018.

[17]. Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. Zhang, “BBDS:Blockchain-based data sharing for electronic medical records in cloud environments,” Information, vol. 8, no. 2, p. 44, 2017.

[18]. X. Zheng, R. R. Mukkamala, R. Vatrapu, and J. Ordieres-Mere, “Blockchain-based personal health data sharing system using cloud storage,” in Proc. IEEE 20th Int. Conf. e-Health Netw., Appl. Services (Healthcom), Sep. 2018, pp. 1–6.

[19]. J. Liu, X. Li, L. Ye, H. Zhang, X. Du, and M. Guizani, “BPDS: Ablockchain based privacy-preserving data sharing for electronic medical records,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2018, pp. 1–6.

[20]. X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, “Integrating blockchain for data sharing and collaboration in mobile healthcare applications,” in Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC), Oct. 2017, pp. 1–5

Cite this article as :

Punugupati Sai Kumar, Dandu Srivamsi, Dr. M. D. Anto Praveena, "A Data Storing and Sharing Solution with Guaranteed Reliability", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 5, pp. 291-300, September-October 2022.

Journal URL : https://ijsrst.com/IJSRST229557