A Certificateless Multi-Receiver Encryption Scheme Based on SM2 Signature Algorithm

M Prasanth, T Syed Ibrahim and K Vanmathi

May 6, 2024

A CERTIFICATELESS MULTI-RECEIVER ENCRYPTION SCHEME BASED ON SM2 SIGNATURE ALGORITHM

ABSTRACT

Provable Data Possession (PDP) allows cloud users to verify data integrity without having to obtain the entire file. All of the existing PDP systems require the Public Key Infrastructure (PKI). The system is flexible, efficient, and enables delegated, private, and public verification. Since it is not attained, 2 Ack-Ib-Dpdp is not sound. Provide an illustration of a general construction that addresses the issue. A new 2 Ack-Ib-Dpdp protocol is obtained by expanding the basic 2 Ack-Ib-Dpdp to several cloud environments. Cloud-based data storage and sharing systems enable users to instantly modify and share content as a group. To ensure that each block inside the shared data can be independently verified as being authentic, group members need to compute signatures for each block. Different users frequently sign different blocks in shared data because to user-performed data modifications. For security concerns, blocks that were previously signed by a user who was kicked out of the group need to be re-signed by a current user. The straightforward method that permits an existing user to download the corresponding portion of the data and re-sign it after user revocation is wasteful given the volume of shared data stored in the cloud.

Key words: Provable Data Possession (Pdp), Cloud Computing, Data Integrity, Public Key Infrastructure (Pki)

1. INTRODUCTION

In the past few years, cloud computing has become more significant in the computer business. It basically makes use of storage devices, computers, and information processing as a service. It relieves the burden of storage management and offers worldwide data access from disparate geographical locations. Additionally, it reduces capital expenditures related to employee maintenance, software, hardware, and other costs. As a result, organisations are paying greater attention to cloud computing. Cloud computing is based on the practice of hiring a third party to do computer operations. It covers the security issues with integrity, confidentiality, and availability of data and services. Convincing cloud customers that their data is retained is crucial because they do not store their data locally. In order to solve this issue, remote data integrity checking is one easy way. Distributed storage and integrity checking are typically necessary when a client stores his data on multiple cloud servers.

1.1 PROVABLE DATA POSSESSION (PDP)

In cloud computing, a cryptographic technique known as Provable Data Possession (PDP) enables users to verify the accuracy of their data without obtaining the entire file. This method typically uses two cryptographic techniques: hashing and digital signatures. PDP stores small bits of data or information alongside the primary data file, allowing users to do integrity checks quickly and effectively from a distance. An essential level of security and confidence to cloud-based data management and storage is added by employing this technique, which allows cloud customers to be assured that their data is safe and undamaged in the cloud.

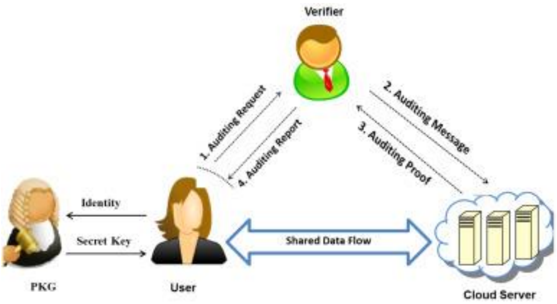

Figure 1 Provable Data Possession

1.2 CLOUD COMPUTING

The pay-as-you-go delivery of computer services via the internet, encompassing servers, storage, databases, networking, software, and more, is referred to as "cloud computing". This method allows users to access and utilise a range of IT services without needing to make a sizable upfront investment in hardware or infrastructure. Cloud computing offers scalability, flexibility, and cost-effectiveness, allowing people and organisations to modify their resource levels in response to demand.

It covers a variety of service models, including Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS), ranging from simple data storage to complex application creation and deployment. Additionally, cloud computing facilitates remote access and collaboration by enabling users to work from any internet-connected location. All things considered, cloud computing has given businesses and consumers unprecedented levels of accessibility, agility, and creativity, utterly revolutionising the IT sector.

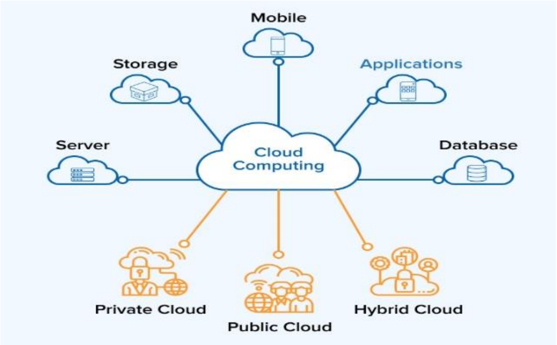

Figure 2 Cloud Computing

1.3 DATA INTEGRITY

To guarantee that it is complete and unaltered, data must be precise, consistent, and dependable throughout its lifecycle. We call this data integrity. In the context of digital information, maintaining data integrity is essential to guaranteeing its dependability and appropriateness for a variety of functions, including analysis, compliance, and decision-making. In order to prevent unauthorised access, corruption, and manipulation, techniques like encryption, checksums, and digital signatures are frequently used to protect data integrity. Data integrity ensures that data is unchanged during processing, transit, or storage, protecting against errors, malicious attacks, and hardware failures. By maintaining data integrity, organisations can rely on it to support essential business operations, adhere to legal requirements, and maintain stakeholder confidence.

1.4 PUBLIC KEY INFRASTRUCTURE (PKI)

A system known as Public Key Infrastructure (PKI) makes it possible to authenticate and communicate securely over unsecure networks like the internet. PKI is based on the fundamentals of asymmetric cryptography, in which every entity has a public key and a private key. The private key is kept secret and is used for decryption and signing, whereas the public key is used for encryption and digital signature verification. By using digital certificates, which are issued by reputable Certificate Authorities (CAs) and link an entity's public key to its identity, PKI promotes trust in digital transactions. By authenticating individuals, devices, and services, these certificates guarantee the confidentiality and impenetrability of data transferred via networks. Virtual private networks (VPNs), secure online browsing (HTTPS), secure email communication, and e-commerce transactions are just a few of the applications in which public key infrastructure (PKI) is essential. In general, PKI offers a strong basis for creating reliable and safe communication networks in the digital era.

2. LITERATURE REVIEW

The idea of Industry 4.0 and its related applications, according to Gang Li [1] et al.'s article, have successfully tapped into newly created technologies including Cyber Physical Systems, Internet of Things, Cloud Computing, and Big Data Analytics. Realising the distributed, cooperative, and automated design and manufacturing workflow is the aim. More specifically, peripherals like software, sensors, and electronics can be combined with cyber-physical systems, the internet of things, and foundational infrastructures to collect and transfer industrial data. Cloud computing techniques facilitate centralised data storage and offer a platform for resource allocation and research that is accelerated and improved for benefits across the industry at the computation infrastructure level. These digital assets have also been organised and valuable information extracted from them through the use of big data analytics, which makes use of the enormous amounts of industrial data gathered from the previously stated phases. The commercial and academic sectors are quite interested in these reasons. Although some pertinent themes have been the focus of special issues of respectable journals, we hope to shed light on the intricate relationship between Industry 4.0 and Big Data in this special issue. We have extended an invitation to five manuscripts for publication following a comprehensive peer review process, thanks to the cooperation of active experts in related fields. Based on several elements of the smart manufacturing scenario, these studies provided surveys, procedures, and frameworks stretching from the infrastructure level to the data processing level and the service level of Industry 4.0. Here, we summarise all of the contributions in this special issue to present a comprehensive view on it.

According to Chunhui Wen [2] et al., this document The banking industry also plans to use big data's cutting edge technology to improve and combine internal and external data related to credit concerns. The big data technology framework has been successfully applied by the Internet of Things. Relying on more efficient machine learning algorithms to provide an accurate credit risk forecast could reduce self-generated losses while increasing earnings for the Internet of Things finance business. This article employs distributed search engine technology to customise web crawlers to extract the required bank card and transaction data from the multi-source heterogeneous data of the Internet of Things financial industry. To preprocess the data, it then creates an inverted table and two Level index files that are used as data sources by big data research systems. Finally, it designs the matching Spark parallel algorithm. Following the identification of the data source, a set of potential indicators and quantification strategies are acquired for the Internet of Things' financial credit risk assessment. Next, the relationship between the risk grading and the indicators is investigated. The scores of several financial business specialists in the industry are integrated with the Mutually Exclusive Collectively Exhaustive (MECE) analysis approach. The random forest algorithm in the big data machine learning library is used to choose the features of the candidate index set. An intelligent early warning model and credit risk assessment are produced by mining the financial risk data from the Internet of Things using a multi-level spatial association rule algorithm based on the hash structure. To achieve its research goals, this study uses 26 Internet of Things financial metrics. Using SPSS26.0 software, it runs sample Kaiser-Meyer-Olkin (KMO) tests and Bartlett sphere tests on the real data and offers a thorough explanation of the factor analysis results. The implementation of the particle swarm method in the random forest parameter optimisation process occurs concurrently with the creation of the Internet of Things' financial credit risk assessment model. The results show that this approach may significantly reduce the chance that banks would make the first and second mistake rates when evaluating the credit risk of financing the Internet of Things. This lessens losses from incorrect credit provision and facilitates the seamless growth of banks' Internet of Things financial operations. This may lead to banks becoming more profitable themselves. Most people are beginning to embrace and value qualities like low cost, great teamwork, etc.

In order to improve the intelligence of the medical system, this study develops and implements a secure medical big data ecosystem on top of the Hadoop big data platform, as suggested by Xiang feng Zhang [3] et al. It was developed in response to the more serious trend that big data security in the medical industry is currently experiencing. In order to enhance patients' comprehension of their treatment state and boost the effectiveness of conventional medical rehabilitation activities, this study proposes a personalised health information system. To guarantee independent data storage, all medical health data is dispersed among numerous different medical institutions. Patients can view their treatment and rehabilitation status from anywhere at any time. As a distributed accounting platform for multi-party maintenance and backup information security, blockchain offers a promising new horizon for medical data sharing innovation. The deployment of a personal health datacenter system using the Hadoop big data platform is explained in this article. The dispersed original data may now be centrally stored and analysed thanks to the data synchronisation module and the independent data gathering system. By utilising the advantages of the Hadoop big data platform, the specially designed health information system for stroke patients has been developed to offer patients personalised health management services and facilitate patient management by medical professionals. The expansion of medical information services is a global trend. Because information technology is advancing so quickly, more and more domestic medical institutions are adopting information-based platforms based on overall construction at a faster rate. These systems have improved the hospital's core competitiveness and service quality. Information technology not only makes medical staff more productive, but it also allows them to see patients more often, builds patient satisfaction and trust, and produces an undetectable scientific image of healthcare facilities. The US health care information industry has made great strides lately. Google and the US Medical Centre work together to give hundreds of millions of people access to electronic health records that doctors can view from a distance.

Xin Huang [4] et al. have proposed that in this system, patients can view examination pictures at any time using a range of electronic techniques, and clinicians use electronic images instead of traditional film for diagnosis in the context of electronic medical data. However, regular access to large amounts of data and storage of electronic data present new challenges. A two-level model combined with medical imaging information is proposed according to the characteristics of medical data with examination as the basic unit; different merging strategies are proposed to improve the storage performance of the files, given the characteristics of the size and quantity of image files generated by various examination types. Since the indexing mechanism solves the problem that SEQ files cannot be read at random without an index, a refined 2Q algorithm is proposed to increase file reading efficiency by caching the read files and the perfected files in separate cache queues, taking into account the time constraints associated with data access. In the experimental comparison, the recommended method performs better in terms of storage and access performance than the baseline strategy. The smart medical component of a smart city is essential. Hospital informatization, intelligence, and electronization have the potential to greatly improve the efficiency of the healthcare system. Massive volumes of electronic medical imaging data present new challenges for the multisource complex data storage and retrieval process, but they also provide robust data support for astute auxiliary diagnostic algorithms. In the new application scenario, doctors use electronic images instead of traditional film for diagnosis, and patients can view the examination photos whenever they want via a range of electronic methods. These questions are frequently based on certain analyses. The medical picture files produced by a certain test should be classified as tiny data based on the quantity and size of the files.

According to Somnath Mazumdar [5] et al., there is an exponential growth in the amount of data that computer systems can currently analyse and use in this system. The enormous amount of data, or so-called "Big Data," created a need for scalable, rapid, and efficient support for existing technologies. New applications and the continuous user assistance that multi-domain computing offered helped facilitate the shift from data-centric to knowledge-centric computing. However, there is still concern about how to move, store, or arrange these enormous data sets throughout data centres (DCs). Because application and DC behaviour (i.e., resources or latencies) are subject to frequent changes, it is especially important to analyse patterns in data access or consumption. The main objective is to find a better data storage site that improves the overall cost of data placement as well as the application's performance (such as throughput). In this survey paper, we offer a state-of-the-art evaluation of cloud-centric big data deployment and data storage strategies. This is an attempt to demonstrate the actual relationship between these two in order to better support Big Data management. We focus on management domains viewed from the perspective of non-functional characteristics. Readers will be able to enjoy the comprehensive analysis of the various technologies associated with Big Data management at the end, and they will receive recommendations for the technologies that best suit their needs for non-functional applications. Furthermore, challenges are posed that highlight the gaps that are there. Over time, applications have evolved from batch, memory-intensive, computational, or batch to streaming and even interactive. As a result, applications are getting longer and more complex. These applications may require frequent access to numerous disparate data sources. During the deployment and provisioning of an application, the user may face several obstacles, such as (i) figuring out where to locate the data and computation in an effective manner and (ii) figuring out how to achieve the required goals while reducing the overall operational expenses of the programme. Data may come from a multitude of sources when the apps are running, including a vast number of Internet of Things-connected devices.

3. RELATED WORK

Safe data delivery requirements may be satisfied by the Multi-receiver Encryption (MRE) approach in both multicast and broadcast situations. In order to comply with rules, China's critical information infrastructure should be secured using Chinese national commercial cryptography techniques. Using Elliptic Curve Cryptography (ECC) to create an MRE scheme is one of the most efficient and adaptable design strategies available today. Conversely, MRE systems based on SM2 elliptic curve public-key encryption are not currently the subject of any research. This work proposes a certificate-less SM2-based multi-receiver encryption technique called CL-SM2-MRE. Under the Random Oracle Model (ROM), we evaluate the security and performance of the CL-SM2-MRE scheme.

4. METHODOLOGY

Data integrity is one of the most important issues with cloud abandon tendencies since there is a lack of identity privacy and customers are unaware of the data auditor across geographically dispersed datacenters. Because of these cloud computing features, there are now a number of worries about user availability, data integrity, and identification. In the end, this affects the proposal of an improved approach to audit data integrity while maintaining identity privacy and facilitating effective user revocation during sharing. We suggest 2 Ack-Ib-Dpdp, a novel multi-cloud authentication protocol with two schemes. Because the basic scheme (2 Ack-Ib-Dpdp) uses an efficient cryptographic primitive called batch signature to support the authentication of multiple data at once, it eliminates correlation between data and offers perfect resilience to data security. It is also efficient in terms of latency, computation, and communication overhead. A group of two Ack-Ib-Dpdp on multiple cloud storage key generation centres (2 Ack-Ib-Dpdp) can produce the keys used in each subgroup in concurrently. Despite the fact that their individual keys are produced by distinct KGCs, all members of the same subgroup are able to compute the same subgroup key. This reduces the issue of focusing the workload on a single entity, which makes it a desirable feature, particularly for large-scale network systems.

5.MODULE DESCRIPTION

5.1 LOGIN & REGISTRATION FOR GROUP MEMBER

In this module, the user first picks a group ID, enters his login and password, and then registers with the Data Cloud Server. This user become a member of that specific group. He then chose his group ID, input his password, and logged in.

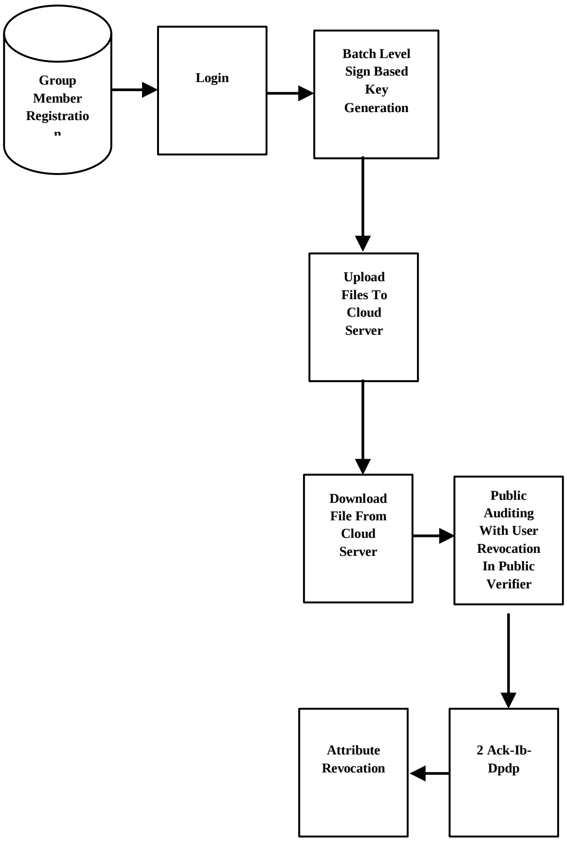
Group Member Registratio n
Login
Batch Level Sign Based Key Generation
Upload Files To Cloud Server
Download File From Cloud Server
Public Auditing With User Revocation In Public Verifier
Attribute Revocation
2 Ack-Ib-Dpdp

Figure 3. Block diagram

5.2 SIGN BASED BATCH LEVEL KEY GENERATION

In the Key Generation module, every group member creates their own public key and private key. After creating a random key, the user produces the public and private keys. Without compromising generality, we assume that user u1 is the original user who produced the shared data. The first user also creates a user list (UL) containing the ids of all the users in the group. The public user list bears the signature of the original user.

5.3 CONFIRM FILES WITH THE CLOUD SERVER.

The user wants to use this module to upload a file. Next, he separated the files into several blocks. Next, he encrypts each block using his public key. Then he develops a signature for each block in order to facilitate authentication. Next, he uploads each block cypher text along with the block id, signature, and signer ID. This metadata and the Key Details are stored by Public Verifier for open auditing.

5.4 GET THE FILE FROM THE CLOUD SERVER

A file in this module is wanted to be downloaded by the next user or group member. After obtaining the secret key, he shares the filename with others. He then entered this code. If this secret key is authentic, the user can decrypt the file that was downloaded. He would have been prohibited by Public Verifier if he had entered the wrong secret key. To verify the validity of this secret key, decode each block and examine the signature. The original file is obtained when all blocks are concatenated and both signatures are equal.

5.5 PUBLIC AUDITING IN PUBLIC VERIFIER WITH USER REVOCATION

The public verifier banned the user who entered the wrong secret key in this module. After that, he deleted the user list and added a public verifier. The Data Cloud Server responds with the prohibited information whenever he tries to download any file. The public verifier is then asked to reverse the revocation by him. The public verifier has finally unrevocated this user. With the included secret key, he was then able to download any file.

5.6 REVOCATION OF ATTITUTE

Our method leverages the idea of proxy re-signatures, whereby the blocks previously signed by a revoked user in the group may be re-signed by the Data Cloud Server using a resigning key. As a result, user revocation efficiency can be significantly raised and current users' computer and communication resources may be easily retained. It is not possible for the Data Cloud Server to sign blocks on behalf of revoked users or existing users at random since it is not in the same trusted domain as each user. Instead, it can only alter the signature of a revoked user to that of an existing user on the same block.

6. RESULT ANALYSIS

The following table displays the accuracy metrics for a number of techniques, including Elliptic Curve Cryptography (ECC) and Ack-Ib-Dpdp. The accuracy rate of Ack-Ib-Dpdp is 99.2%, significantly higher than that of ECC, which is 78%. Elliptic Curve Cryptography is a widely used cryptographic method that is renowned for its effectiveness and security, particularly in environments with constrained resources. However, because of its decreased accuracy, there may be limitations in some use cases or datasets. But Ack-Ib-Dpdp, or "Acknowledgment-based Incremental Block-level Dynamic Provable Data Possession," achieves a far higher accuracy rate, proving its effectiveness in verifying data integrity in cloud storage systems. Since this algorithm most likely employs state-of-the-art techniques to ensure resilience and reliability in data verification processes, it is a good choice for secure data management in cloud environments. More investigation and testing could be necessary to identify the precise reasons for the varying accuracy levels displayed by these algorithms as well as to identify potential areas for optimisation or improvement.

algorithm	**accuracy**
ECC	78
Ack-Ib-Dpdp	99.2

Table 1. Comparison table

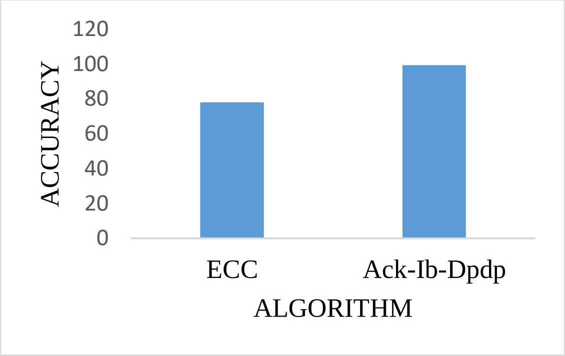

Table 2. Comparison graph

6. CONCLUSION

In conclusion, building a robust system for public auditing with user revocation in a multi-user environment requires thorough examination of a number of modules, including user registration, key generation, file management, and revocation procedures. If the system includes user-friendly input and output designs, undergoes thorough system testing, and ensures that every aspect is meticulously attended to during the implementation process, it may successfully satisfy user requests while safeguarding data integrity and security. Furthermore, continuous monitoring, maintenance, and adaptation to shifting technological advancements and user needs are necessary to sustain the system's effectiveness and relevance over time. In cloud-based data storage and sharing scenarios, a well-executed system deployment supported by thorough testing and continuous improvement efforts may eventually lead to increased user satisfaction, increased productivity, and strengthened security posture.

7. FUTURE WORK

When considering future development, a variety of strategies may be looked into to enhance the system's functionality, scalability, and security. One area of focus may be on redesigning the user interface to make it more approachable and user-friendly, and possibly including technologies like multi-factor authentication and biometrics for added security. Further research and development efforts could also be directed towards improving the functionality of the system, particularly with regard to handling large data sets and increasing the throughput of file uploads and downloads. Integration with cutting-edge technologies like block chains for decentralised authentication and data verification should be looked at in order to improve system trust and transparency. Additionally, more research into advancements in cryptography and related domains might lead to the development of safer and more efficient techniques for attribute and user revocation. Collaborating with academic establishments and industry players could facilitate the exchange of best practices and knowledge, enabling the system to adapt and grow over time to the ever-changing needs and challenges of cloud-based data sharing and storage environments.

8. REFERENCES

1. Innovations in big data and Industry 4.0, Enterprise Information Systems, 13 (2) (2019) 145–1. G. Li, J. Tan, S. S. Chaudhry.

2. Computer Systems of the Future, 124 (6) (2021) 295–307, Big data driven internet of things for credit evaluation and early warning in finance, C. Wen, J. Yang, L. Gan, Y. Pan.

3. Research on a Hadoop and blockchain-based intelligent medical big data system EURASIP Journal on Wireless Communications and Networking, 2021 (1), 1–21; Zhang, X., Wang, Y.

4. Engineering Mathematical Problems 20. A medical image storage and retrieval method for examination series based on Hadoop.

5. A survey on data placement and storage strategies for the cloud big data ecosystem that was published in the Journal of Big Data.

6. W. Rajeh, Hadoop distributed file system security issues and an examination of the issue of unauthorised access, Journal of Information Security, 13 (2) (2022) 23–42
1-2 (4) (2019) 1–8.

7. Array. Big data: Hadoop framework flaws, security issues, and assaults, G. S. Bhathal, A. Singh.

8. Kapil, A. Agrawal, R. A. Khan, Big data security challenges: Hadoop viewpoint, International Journal of Pure and Applied Mathematics, 120 (6) (2020), 11767–11784.

9. B. H. Husain, S. R. Zeebaree, et al., "Improvised distributions framework of hadoop: A review," International Journal of Science and Business, 5 (2) (2021), 31–41.

10. A thorough literature study by M. Naisuty, A. N. Hidayanto, N. C. Harahap, A. Rosyiq, and G. M. S. Hartono, Data security on Hadoop distributed file system by using encryption techniques, Journal of Physics Conference Series, 1444 (4) (2020) 1–8.