Journal of IoT in Social, Mobile, Analytics, and Cloud (ISSN: 2582-1369)
www.irojournals.com/iroismac/

Data Immutability Challenges: A Security Analysis of Digital Archiving Systems

T. Senthilkumar[1], S. Rajasekaran[2]

[1]Assistant Professor, Department of Electrical and Electronics Engineering, RVS College of Engineering and Technology, Coimbatore, India
[2]Professor, Department of Electrical and Electronics Engineering, Vignana Bharati Institute of Technology, Hyderabad, India

E-mail: [1]texrosen@gmail.com, [2]rajasekaran@vbithyd.ac.in

Abstract

This study examines how well 'cloud computing' and 'digital archive resource management' work in the present time and what advantages they provide. Additionally, it explains a novel approach to using cloud computing by means of blockchain with an appropriate technique for sharing and distributing digital archive contents. The combination of virtualization and cloud storage technologies results in a widely accessible dataset. Specifically, it demonstrates its high security, excellent scalability, and ideality for standardization. This article also investigates the feasibility of implementing the archival bond in a blockchain-based record-keeping infrastructure. To achieve this, it suggests a database backend that preserves the immutability of the blockchain while facilitating metadata updates and improved search capabilities.

Keywords: Archive information, security, big data, block chain, confidentiality management, digital archives

1. Introduction

Integration and exchange of digital archive resources has been a topic of interest recently. However, as library informationization levels rise, the widespread use of digital building techniques presents serious threats to the safety of digital archives at universities and colleges. Therefore, this is the number one issue with creating university libraries that has to be addressed right now.

1.1 Acceptance of the Open Archival Information System (OAIS)

Having a reliable digital repository ensures that the whole repository system is compliant with the OAIS Reference Model. For digital archiving services to be successful, all

DOI: https://doi.org/10.36548/jismac.2022.3.007
Received: 17.08.2022, received in revised form: 02.09.2022, accepted: 19.09.2022, published: 30.09.2022

relevant parties must be on the same page about their goals and the steps to get there. The Reference Model provides a standard vocabulary and ideas for comparing digital archive architecture and operations. In addition, the OAIS offers a functional model i.e., the precise functions carried out by the repository, such as storage or access, and an accompanying information model for generating metadata in service of preservation and accessibility through time. Digital repository developers should make time to learn these concepts and check that their code follows them [1-4].

1.2 Duty of Administration

A reliable digital archive will show that it is firmly committed to adopting the many best practices and standards established by its industry. This is especially true for the standards that have a direct bearing on the archive's long-term health. Those in administrative roles are tasked with ensuring the infrastructure, disaster recovery plans, and security measures, wherein all conform to be applicable to national and international standards. The trustworthy repository will regularly gather and communicate data measurement results with depositors, and it will consistently meet or exceed community performance criteria. On a regular basis, it will bring in community experts from the outside to validate and/or certify its processes and procedures. All terms relating to the receipt, custody, use, and withdrawal of deposits should be set out in a written agreement with depositors. In addition, the business will routinely include risk management and contingency planning into its yearly strategic planning [5,6].

1.3 Continuity of Operations

Digital repositories that want to prove their reliability will set up methods that prove their longevity. On behalf of depositors and users, their stated goals will emphasize preservation, administration, and continued access to digital cultural goods. A legal identity and standing commensurate with the breadth of their duties will be conferred upon them. They will be honest and open in their dealings with customers. The appropriate number of employees and their specific areas of expertise will be employed, and workers will be encouraged to pursue chances for continuing education, such as by attending and presenting at conferences, to keep their skills up to date. To allow for suitable expansion, the repository's rules and procedures will be reviewed on a regular basis, and any new processes or procedures will be thoroughly evaluated for scalability. A detailed succession plan or escrow arrangements will be created in conjunction with community experts, depositors, and other

comparable parties that details all pertinent material and identities of trustworthy inheritors in the event that the repository ceases operations [7,8].

1.4 Consistency with Budgetary Resources

A reliable online archive should be able to demonstrate its long-term viability. If they want to be considered reliable, repositories must follow all standard best practices and operate with a long-term strategy. At least once a year, the health of the company and its finances should be evaluated. A general ledger approach should be used. Risk, benefit, investment, and spending must be consistently balanced across both the short- and long-term economic planning cycles. Both the operating budget and the reserve funds need to be sufficient [9]. Figure 1 shows the data archiving system.

Figure 1. Data Archiving System [26]

1.5 Appropriateness of Methods and Technologies

There is a wide variety of preservation measures that are now being advocated by community specialists. A reliable digital archive will take into account all relevant possibilities and share freely about the appropriateness of different approaches. Plus, it will make sure that it has all the necessary hardware set up. Inventory management software is used to track and manage stock levels and the many ways in which items may be acquired, stored, and retrieved. Policies and strategies for upgrading existing infrastructure will also be included in the central database, when it becomes necessary. The repository will follow all applicable guidelines, and the team will be well-equipped to comprehend and put them into effect. Additionally, the system's individual parts and overall operation will be subjected to periodic third-party audits [10].

1.6 Intent Model Reference

To help those who aren't necessarily specialists in simulation technology, software engineering, or IT, support services become more familiar with the capabilities and

difficulties related to maintaining, supporting, and deploying M&S systems. This study may be useful in developing to-do lists for addressing potential issues with newly acquired DHS simulation software. This study is hoped to serve as a starting point for productive dialogue among DHS staff, external developers, and the wider M&S community. The report's purpose is to assist those whose work involves installing, supporting, and maintaining homeland security simulations and models. Those working on models and simulations could be experts in certain areas but unaware of others while having a thorough understanding of the nuances involved, because of the breadth and depth of M&S at DHS. The most important aspects of M&S for homeland security from a bird's eye perspective has been covered in this paper. Its goal is to enlighten workers about a variety of problems and offer answers to pressing concerns in order to facilitate their work in areas with which they may be unfamiliar [11,12]. Figure 2 shows the concept of Digital Archiving.

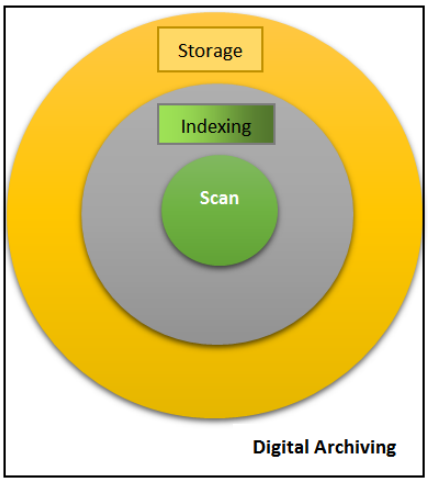

Figure 2. Simple concept of Digital Archiving

1.7 Motivation

University digital archives have entered a new phase of development because of the rapid pace of office automation and other technological breakthroughs. When building a fully digital campus, digitizing university records is a must and provides a useful indicator of how far along that process is. Yet, as archive information technology advances, university digital archives face increasing challenges in data security.

2. Literature Survey

Unique solutions are accomplished for a few of the systems under scrutiny. As a result, they weren't considered relevant to the issue at hand in this work. What follows is a

quick explanation of each of them. Galiev et al. [13] mentioned that ARCHAIN is based on an existing archive system. While the authors did shed light on the difficulties of archival design integration, no practical solution was proposed for use with TrustChain. Nonetheless, the article proved the premise of the initial TrustChain concept that blockchains can and will be used to guarantee the authenticity of material stored in long-term archives.

While the Cliegon E-Archive system's proposed solution is intriguing, it doesn't seem practical for the TrustChain solution, since it relies on Interplanetary File System (IPFS), a distributed file storage system [14]. TrustChain cannot rely on IPFS as a data storage solution since its files must be searchable; thus, the company must create its own database management system. Other academics, however, have looked at the idea of combining IPFS and blockchain together to guarantee data accuracy; for instance, Naz et al. [15] proposed a similar solution. While such a system might be useful for archiving historical documents, it is not the focus of this study. The TrustChain solution is.

2.1 Exposition of the Problem

To reconcile the seemingly incompatible goals of immutable blockchain technology with long-term data storage, this article proposes a TrustChain-based data infrastructure. The database works in tandem with the blockchain to facilitate indexing and searching, as well as to provide a means by which TrustChain users can update certain metadata information. However, a purely immutable data structure is insufficient for the system to do what it sets out to do; instead, a dual storage system consisting of a trustworthy blockchain at its core to ensure data integrity and a somewhat modifiable supporting system is required. Comparing existing specialist archive and general database systems that use blockchains to guarantee data integrity, a literature analysis is performed to establish the best suitable technology for the proposed supported system. By doing a literature study, it could be ascertained which underlying technology solutions are most often used and how well they match with the proposed system.

3. Information Security for File Management System

The information security issues facing university digital archives have grown increasingly pressing as the level of informationization of archives has increased. On one hand, the seriousness and severity of the security issue is not recognized. It might, for instance, ignore website security upkeep or system installation in favour of focusing on

website creation or computer performance. One the other hand, managers' lack of security awareness leads to a lack of focus on information security, which in turn leads to issues with the development and upkeep of security software. Therefore, raising awareness of information security among digital archive administrators is the first step toward bettering the security of information handled in digital archives [16-19].

3.1 Concerns about malware in the network environment and their potential impact on digital data

The constant improvement of network technology has sped up the development of information applications. As a result of the Internet's accessibility, harmful programmes like viruses and trojans pose serious challenges to the information security of many networks. The security of data stored in digital archives is similarly vulnerable to the proliferation of dangerous software through the Internet. Statistics show that in 2009 alone, Trojans took control of as many as 262,000 domestic host IP addresses. The archival information systems of universities are under constant assault from Trojans.

3.2 Protecting System

A reliable digital archive will have security measures built into every aspect of its infrastructure. Community standards, such as those for data duplication, redundancy, authentication, firewalls, and backups, shall be adhered to. Staff will be well trained, and the repository will have established policies and strategies for disaster prevention, mitigation, and recovery. There will be an emphasis on procedures that deal with data integrity in order to protect data from corruption, monitor data changes, and recover damaged data. The depositor will be informed of the discrepancies and any actions taken, and any modifications such as deletions, corruptions, or repairs, recorded [20].

3.3 Responsibilization in Procedure

Because of the interconnected nature of the duties and operations it performs, a reliable digital repository will be held to all applicable standards and regulations. Any interested party will have access to the repository's procedures manual. Metrics and checks will be set up to make sure everything keeps running well. Actions made for the sake of preservation (such as migration, emulation, etc.) shall be documented and explained considering generally accepted standards. As the repository's requirements, third-party

service providers, and authorised communities evolve, there will be opportunities to offer input and discuss the implications [21].

3.4 Skill in digital file management

Many individuals who deal with computers and handle files at today's institutions are not very adept at digital file management. A full-text, multimedia, and catalogue database of archives is necessary for digital file management. However, modern university archivists are not well-versed in the tools, technologies, and processes involved in computer and network operations. Due to this, digital libraries are slowed in their ability to acquire new data [22].

3.5 Institutional countermeasures for shoring up the data security of digital archives

There are a growing number of potential points of failure in data security, including the storage medium, the network infrastructure, the hardware components, and the security awareness of end users. Therefore, it is crucial to continue strengthening the building among those in charge of managing digital archives, establish a system to guarantee information security technology for digital libraries, and so on.

3.6 Improve knowledge of data protection among those in charge of digital files

Enhancing the security of digital libraries is predicated on raising the understanding of digital archive administrators about the need of data protection. If file managers aren't made aware of the need of data security, no precautions will be taken to keep data in files secure. Therefore, there are a variety of options for raising file management staff's awareness of information security, including specialised information security lectures, case studies, business research, etc. People in charge of file management will be better able to focus on data protection because of this. As a result, people will be more aware of the need of keeping their personal information safe online, which will reduce the likelihood of data breaches and other breaches of privacy [23].

3.7 Creating a safe and secure technological infrastructure

There are major practical advantages to the availability of digital archives that allow for distant sharing, yet some individuals or businesses may unlawfully get some vital data or information obtained over the network. As a result, more resources must be put into developing and deploying network security technology, and also a network technology guarantee system needs to be set up [24].

3.8 Recruiting top-tier talent in the field of information security

While current workers at higher education institutions charged with overseeing digital files get little in the way of ongoing training, their rudimentary knowledge of information security management may benefit from it. As a result, the appropriate units within the institution should boost their spending on infrastructure and develop more advanced training programmes, so that managers can do their jobs more competently. The strategy should include up-to-date management data, and digital technological data.

3.9 Model for a System that can support Trust in a Blockchain

Both the underlying technology and the metadata model had to be decided upon before the supporting system could be modelled. The searchability is not a technical challenge, although indexing is accomplished mostly by the underlying technology of the supporting system. Since archive searches are conducted via exploration of information collections, choosing the right metadata standard is critical for the supporting system [25].

3.10 Examining the Safety of the System

The cloud server is often portrayed as trustworthy yet inquisitive, meaning that it understands and adheres to the protocol requirements for cloud services but is nevertheless interested in learning more about archive files and data. That is to say, people do not have faith in the cloud server. The security of the confidentiality management model of cloud-based digital archives is theoretically examined, including the security of archive files, archive data, and archive feature data; that is, the likelihood that the untrusted cloud server obtains sensitive archive information in accordance with the encrypted archive files, encrypted archive data, and encrypted archive feature data submitted by the local server is examined.

4. Conclusion

Managing archival material in higher education institutions is moving in the direction of digital archives. Securing digital archives at universities is a complex and time-consuming undertaking that calls for the integration of several disciplines and resources, including software, regulations, organisational structures, mindsets, and human resources. The number of potential threats to the privacy of data stored in digital archives is growing in tandem with the expansion of both social media and scientific knowledge. All university archives should

place a high priority on working together to increase the efficiency of university archives administration. This research will continue to investigate the following issues in the future:

- Ways to simplify the archival release mechanism and the archival search paradigm to reduce load on the local server.
- How to optimize efficiency and safety by developing a variety of feature-construction strategies for use with archival data.
- How the suggested approach may be implemented in a cloud-based digital archive management system.

References

[1] J. Wen: Study on the Construction of Digital Archives Based on Cloud Computing Journal of archives and construction, Vol. 5 (2011) No.1, p.46.

[2] DinkarSitaram and GeethaManjunath. Migrating to the Cloud: Developing Applications in the New World of Cloud Computing .Defense Industry Press, 2015.

[3] Gao Binsheng. Countermeasures and Measures for Security Protection of Archives Information System in Colleges and Universities. China Management Information, 2015, 18(23): 169-170.

[4] Yan, W.; Li, G.; Wu, Z.; Wang, S.; Yu, P. Extracting diverse-shapelets for early classification on time series. World Wide Web 2020, 23, 3055–3081.

[5] Goyal, Palash, Sujit Rokka Chhetri, and Arquimedes Canedo. "dyngraph2vec: Capturing network dynamics using dynamic graph representation learning." Knowledge-Based Systems 187 (2020): 104816.

[6] Bai, B.; Li, G.; Wang, S.; Wu, Z.; Yan, W. Time series classification based on multi-feature dictionary representation and ensemble learning. Expert Syst. Appl. 2021, 169, 114162.

[7] PREMIS Editorial Committee. PREMIS Data Dictionary for Preservation Metadata. 2015.

[8] Bandara, E.; Ng, W.K.; de Zoysa, K.; Fernando, N.; Tharaka, S.; Maurakirinathan, P.; Jayasuriya, N. Mystiko—blockchain meets big data. In Proceedings of the 2018 IEEE International Conference on Big Data, Seattle, WA, USA, 10–13 December 2018; pp. 3024–3032.

[9] Pan, J.; Zhang, C.; Wang, H.; Wu, Z. A comparative study of Chinese named entity recognition with different segment representations. Appl. Intell. 2022, 1–13.

[10] Xu, G.; Wu, Z.; Li, G.; Chen, E. Improving contextual advertising matching by using Wikipedia thesaurus knowledge. Knowl. Inf. Syst. 2015, 43, 599–631.

[11] Liu, Jin, Yihe Yang, and Huihua He. "Multi-level semantic representation enhancement network for relationship extraction." Neurocomputing 403 (2020): 282-293.

[12] Permatasari, I.; Essaid, M.; Kim, H.; Ju, H. Blockchain Implementation to Verify Archives Integrity on Cilegon E-Archive. Appl. Sci. 2020, 10, 2621.

[13] Galiev, A.; Prokopyev, N.; Ishmukhametov, S.; Stolov, E.; Latypov, R.; Vlasov, I. Archain: A novel blockchain based archival system. In Proceedings of the 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability, London, UK, 30–31 October 2018. Available online:World Wide Web 2021, 24, 25–49.

[14] Bandara, E.; Liang, X.; Shetty, S.; Ng, W.K.; Foytik, P.; Ranasinghe, N.; Zoysa, K.D.; Langöy, B.; Larsson, D. Lekana-Blockchain Based Archive Storage for Large-Scale Cloud Systems. In Proceedings of the International Conference on Blockchain, Honolulu, HI, USA, 18–20 September 2020.

[15] Naz, M.; Al-zahrani, F.A.; Khalid, R.; Javaid, N.; Qamar, A.M.; Afzal, M.K.; Shafifiq, M. A secure data sharing platform using blockchain and interplanetary file system. Sustainability 2019, 11, 7054.

[16] McConaghy, T.; Marques, R.; Müller, A.; Jonghe, D.D.; McConaghy, T.; McMullen, G.; Henderson, R.; Bellemare, S.; Granzotto, A. Bigchaindb: A Scalable Blockchain Database. Available online: https://gamma.bigchaindb.com/whitepaper/bigchaindb whitepaper.pdf

[17] Muzammal, M.; Qu, Q.; Nasrulin, B. Renovating blockchain with distributed databases: An open source system. Future Gener. Comput. Syst. 2019, 90, 105–117.

[18] Helmer, S.; Roggia, M.; el Ioini, N.; Pahl, C. Ethernitydb–integrating database functionality into a blockchain. In Proceedings of the European Conference on Advances in Databases and Information Systems, Budapest, Hungary, 2–5 September 2018; pp. 1–8.

[19] Batista, D.; Kim, H.; Lemieux, V.L.; Stanˇci´c, H.; Unnithan, C. Blockchains and Provenance: How a Technical System for Tracing Origins, Ownership and Authenticity Can Transform Social Trust. In Building Decentralized Trust: Multidisciplinary Perspectives on the Design of Blockchains and Distributed Ledgers; Lemieux, V.L., Feng, C., Eds.; Springer: Cham, Switzerland, 2021; pp. 111–128.

[20] Brali´c, V.; Kuleš, M.; Stanˇci´c, H. A model for long-term preservation of digital signature validity: TrustChain. In Integrating ICT in Society; Atanassova, I., Zaghouani,

W., Kragi´c, B., Aas, K., Stanˇci´c, H., Seljan, S., Eds.; University of Zagreb: Zagreb, Croatia, 2017; pp. 89–113.

[21] Chen, L. Road vehicle recognition algorithm in safety assistant driving based on artifificial intelligence. Soft Comput. 2021, 1–10.

[22] Wang, Hongbing, and Jian Zhang. "Security Research in Personnel Electronic File Management Based on Blockchain Technology." Security and Communication Networks 2022 (2022).

[23] Habibzadeh, Hadi, Brian H. Nussbaum, Fazel Anjomshoa, Burak Kantarci, and Tolga Soyata. "A survey on cybersecurity, data privacy, and policy issues in cyber-physical system deployments in smart cities." Sustainable Cities and Society 50 (2019): 101660.

[24] Nagasubramanian, Gayathri, Rakesh Kumar Sakthivel, Rizwan Patan, Amir H. Gandomi, Muthuramalingam Sankayya, and Balamurugan Balusamy. "Securing e-health records using keyless signature infrastructure blockchain technology in the cloud." Neural Computing and Applications 32, no. 3 (2020): 639-647.

[25] Völter, Fabiane, Nils Urbach, and Julian Padget. "Trusting the trust machine: Evaluating trust signals of blockchain applications." International Journal of Information Management (2021): 102429.

[26] Website: https://mapsystemsindia.com/resources/uses-of-digital-archiving.html

Author's biography

T. Senthilkumar is currently working as an Assistant Professor in the Department of Electrical and Electronics Engineering RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India. His area of research includes electrical machines, optimization techniques and FPGA.

S. Rajasekaran is currently working as a Professor in the Department of Electrical and Electronics Engineering, Vignana Bharati Institute of Technology, Hyderabad, India. His area of research includes electric drives, controller design, Internet of Things and FPGA.