Unlocking Data Potential with DMChain: Blockchain Framework for Sharing Datasets

Mohamed A. Dini, Simeon O. Ajakwe, Saviour I. Igboanusi, Dong-Seong Kim, Jae Min Lee, Taesoo Jun
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
(m.raji202010, simeonajlove, ikechisaviour)@gmail.com (dskim, ljmpaul, taesoo.jun)@kumoh.ac.kr

***Abstract*—In the realm of dataset sharing, ensuring secure and swift dataset delivery stands as paramount. This study introduces DMChain, a blockchain-based dataset marketplace, which not only facilitates efficient and dependable transactions but also empowers data owners by affording ownership rights and monetization avenues to users. The DMChain framework, enabled by blockchain technology, guarantees both rapid and secure data sharing. The system's implementation, carried out using Ethereum's Georli testnet and employing smart contracts, prioritizes privacy and security. Comparative analysis against related works confirms DMChain's potential as a promising solution to the challenges in dataset sharing. It offers a decentralized, transparent, and robust network, underpinning rapid, secure, and transparent data sharing.**

***Index Terms*—5G, Blockchain, Data, Datasets, Ethereum, Marketplace, Security, Smart contracts.**

I. Introduction

Datasets are now essential to research in a variety of sectors since they provide essential resources for data-driven analysis and innovation. Their ability to provide insightful information and encourage progress accounts for their popularity [1]. Dataset sharing does face difficulties, despite its importance. Open sharing can be hampered by sensitive data, legal restrictions, privacy concerns, and the possibility of abuse. To ensure quality and reproducibility, datasets frequently need careful curation and documentation [2]. The sharing landscape is further complicated by the need to build trust among contributors, worries about intellectual property, and a lack of standardized sharing channels.

Centralized platforms and peer-to-peer exchanges are two common ways to share datasets. While centralized systems are convenient, they frequently control access, limiting the autonomy of data owners and stifling innovation. Moreover, because the data is under the authority of a single company, they can present privacy and security issues [3]. Ultimately, a lack of trust in the quality and source of a dataset due to exposure to unauthorized access in the centralized platform affects the overall output of any system that uses such dataset as input [4]. On the other hand, peer-to-peer sharing offers direct contacts but lacks defined standards, making discovery and building trust difficult. Peer-to-peer sharing may also not be appropriate for huge datasets owing to bandwidth limitations. Both methods have problems with data quality verification, building trust, and potential exploitation, which emphasizes the need for improved sharing channels [5].

Blockchain emerges as a promising solution to the drawbacks seen in traditional dataset-sharing methods. In the realm of centralized platforms, blockchain's decentralized architecture offers a remedy to issues of control and privacy, as data owners retain authority over their datasets while ensuring data integrity through immutable records [6], [7]. Similarly, in peer-to-peer exchanges, blockchain's trust-establishing capabilities and standardized protocols address challenges related to data quality assurance, trust establishment, and data discovery. Additionally, blockchain's encryption features enhance security and privacy, countering potential exploitation concerns. By leveraging these advantages, blockchain offers a comprehensive approach to overcoming the limitations inherent in current dataset-sharing paradigms [6], [8].

This study's contributions revolve around leveraging blockchain technology to establish an efficient and secure data-sharing framework. This approach facilitates rapid and dependable data transactions while empowering data owners with ownership rights and avenues for dataset monetization. The rest of the paper is organized as Section II for system design and experimental setup, Section III for result discussion and model evaluation, and Section IV to conclude the study.

II. Proposed DMChain System

Users can search, browse, and request access to datasets using these platforms. Examples include data-sharing websites, online databases, and repositories managed by organizations. Despite their convenience, centralized methods raise concerns about data privacy, control, and ownership, along with limited interoperability between platforms. Relying on a central authority can also introduce bottlenecks, governance problems, and potential security risks [9], [10].

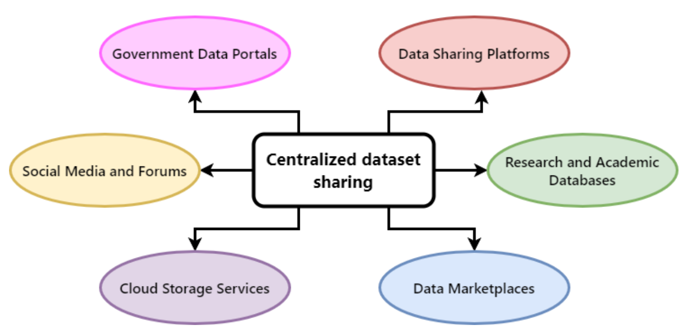

Fig. 1. Centralized Dataset Sharing Methods

Fig. 1 depicts several methods that have been utilized for centralized dataset sharing. Typically, those who create datasets submit their data to centralized platforms, making it available for use by others, particularly researchers. These datasets can be either public or private, with privacy issues arising from sensitivity, regulations, expenses, and security considerations. Data owners utilize physical security measures, encrypted storage, and data-sharing agreements to address these issues. Sharing confidential datasets requires both trust and ethical usage. For public datasets, users frequently seek out direct website or link access when wanting to obtain a particular dataset. Otherwise, they must ask the owners for access, accept their rules, and wait for permission for private datasets.

Fig. 2 depicts the system's composition, featuring interconnected components. A blockchain foundation ensures secure and transparent data transactions, with smart contracts governing user interactions. Users participate in dataset sharing and marketplace actions, enabled by these contract functionalities. Furthermore, the system employs IPFS (InterPlanetary File System) for decentralized and efficient dataset metadata storage and retrieval. This collaboration establishes a unified ecosystem, delivering secure and efficient dataset sharing within a tamper-resistant decentralized framework.

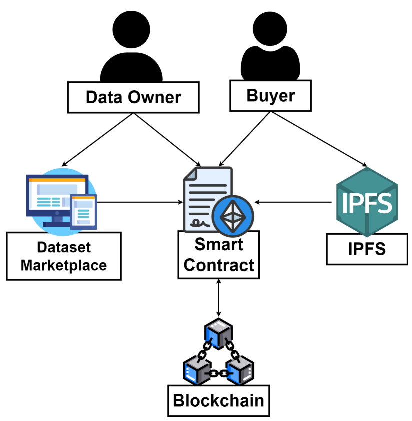

Fig. 2. Our Proposed Dataset Marketplace System Flow

The proposed DMChain system is optimized for different operational elements by the smart contract features built in. Users are given the ability to register, create dataset listings, make and accept offers, complete transactions, and access transaction history. Users are able to efficiently trade datasets thanks to the smooth experience created by these functions. Users who have registered can create listings with important information including dataset type, price, and duration, browse the listings that are already available, make and accept offers, and ultimately complete transactions through fund transfers. Users can also check transaction statuses, see submitted offers, and access their buy and upload history. These features improve the user experience holistically, guaranteeing efficient and effective participation in the dataset market.

Each smart contract function works independently to expand the operational range of the dataset marketplace. User registration is made possible by the $register()$ function, while new dataset listing generation is made possible by createListing(). Users can investigate listing details using the **viewListing()** and **viewAllListings()** functions. Users can develop and examine offers with the help of the **createOffer()** and **viewSubmittedOffers()** functions. The **acceptOffer()** function is used to accept an offer, while the **finalizeTransaction()** function closes the transaction. Access to previous activities is provided by additional procedures like **getUserPurchaseHistory()** and **getUserUploadHistory()**. The functions **getTransactionStatus()** and **viewAvailableOffers()** allow users to browse the offers that are currently available. Together, these features provide a strong, user-centric environment that enables frictionless dataset sharing and exchange.

The blockchain architecture for the proposed DMChain enables decentralized dataset trading using smart contracts on a Georli test network powered by Ethereum and carried out on Remix IDE. The Solidity programming language was employed in the creation of smart contracts. The MSI computer runs Windows 11 with an Intel(R) Core(TM) i5-8500 CPU clocked at 3.00GHz, 6 Core(s), NVIDIA GeForce GT 1030, GPU CUDA: 0 (Tesla K80, 11441.1875MB), and 36GB RAM was used to test Truffle. Gas prices and security analysis of smart contracts are used to validate the proposed system.

III. RESULT DISCUSSION

The findings in Table I validate the computational resources required for executing transactions and achieving completion in the proposed system. These resources are associated with different smart contract deployment operations that result in modifications to the smart contract information.

TABLE I
GAS COST ANALYSIS FOR ALGORITHMS AND FUNCTIONS IN THE SMART CONTRACTS

Function	Gas
Register	62187
getPurchaseHistory	0
getUploadHistory	0
createListing	354,948
viewListing	0
viewAllListing	0
nextListingId	0
createOffer	146140
accepteOffer	62104
finalizeTransaction	62104
viewOffer	0
viewAvailableOffers	0
viewSubmittedOffers	0
nextOfferId	0

Gas cost analysis served as a means to evaluate the smart contract's performance. When accessed from a full node, the smart contract's results incurred no gas costs for read functions, rendering a gas value of zero. Beyond user registration and profile adjustments, DMChain encompasses a range of operational capabilities. These encompass tasks such as adding and generating dataset listings, facilitating offers

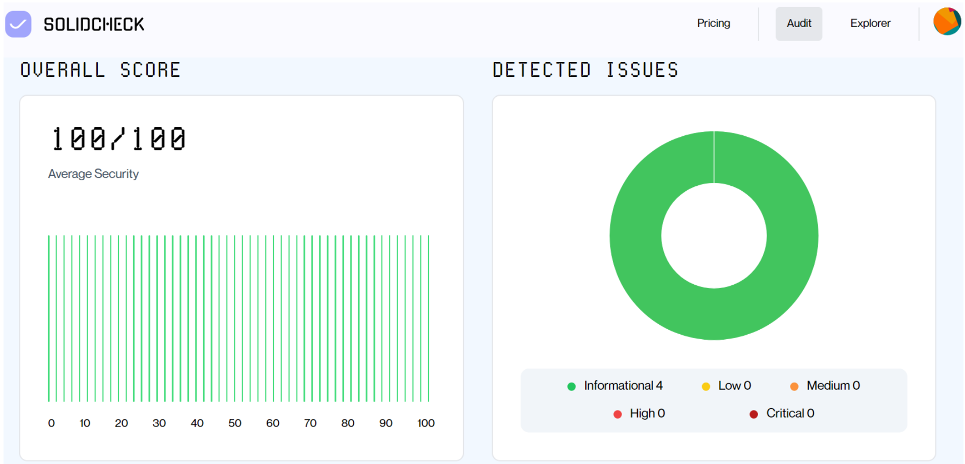

Fig. 3. The smart contract underwent a security evaluation using SolidCheck, demonstrating its robustness and reliability.

and their acceptance, and ultimately concluding transactions. Furthermore, the smart contract security analysis conducted using SolidCheck yielded highly favorable results, with an average security score of 100% as can be seen in Fig. 3. This comprehensive assessment revealed the absence of critical or high vulnerabilities within the examined smart contracts, reinforcing the robustness of the system's design. While the analysis did detect four informational errors, their negligible impact and minimal relevance to the overall security posture suggest that they pose no substantial threat. This impressive security assessment underscores the effectiveness of the system's security measures and highlights its suitability for secure data transactions within the proposed blockchain ecosystem.

IV. CONCLUSION

In response to the pervasive use of datasets and the challenges associated with conventional sharing methods, this study presents DMChain, a blockchain-based dataset marketplace, as an innovative solution. DMChain leverages the transparency and security of blockchain technology, streamlining interactions through smart contracts to address issues like privacy, trust, and efficiency. The framework allows data owners and users to engage in secure and tamper-resistant transactions while maintaining control over access and ownership. The system was deployed and tested using Remix IDE on private and public networks. Looking ahead, integrating Non-Fungible Tokens (NFTs) and developing a user-friendly Decentralized Application (DApp) interface hold the potential to elevate this platform's impact, transforming it into a pioneering ecosystem for data exchange across diverse domains.

V. ACKNOWLEDGEMENT

This research received support from multiple sources: the National Research Foundation of Korea (NRF) through the Ministry of Education, Science, and Technology (MEST) under grant number 2018R1A6A1A03024003, the Ministry of Science and ICT (MSIT) through the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-2020-0-01612) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and the Korea Ministry of SMEs and Startups in 2022 under project S3311338 for Industry-University-Research Institute platform cooperation R&D.

REFERENCES

[1] M. S. Gadelrab and A. Ghorbani, "A new framework for publishing and sharing network and security datasets," in *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, 2012, pp. 539–546.

[2] S. O. Ajakwe, D.-S. Kim, and J. M. Lee, "Drone transportation system: Systematic review of security dynamics for smart mobility," *IEEE Internet of Things Journal*, 2023.

[3] Y. Pan, Z. Cheng, Y. Liu, B. Wang, W. Wang, and C. Zhu, "Research on trusted data sharing method for multi-party edge terminals," in *2022 IEEE International Conference on Unmanned Systems (ICUS)*, 2022, pp. 204–208.

[4] S. O. Ajakwe, V. U. Ihekoronye, D.-S. Kim, and J. M. Lee, "Pervasive intrusion detection scheme to mitigate sensor attacks on uav networks," *Korean Institute of Communication and Information Sciences (KICS) summer conference*, pp. 1267–1268, 2022.

[5] J. Liu, Y. Zhu, H. Wang, J. Guo, and M. Guan, "Design and implementation of a secure peer-to-peer file transfer system based on ip multicast and diffie-hellman algorithm," in *2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 2022, pp. 520–524.

[6] Z. Wu, H. Pan, X. Si, H. Qian, and F. Cai, "A fair and reliable data trading scheme based on blockchain," in *2022 2nd International Conference on Computer Science and Blockchain (CCSB)*, 2022, pp. 34–41.

[7] S. O. Ajakwe, I. I. Saviour, J.-H. Kim, D.-S. Kim, and J. M. Lee, "Banda: A novel blockchain-assisted network for drone authentication," in *2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2023, pp. 120–125.

[8] Z. Cheng, Y. Pan, Y. Liu, B. Wang, X. Deng, and C. Zhu, "Vflchain: Blockchain-enabled vertical federated learning for edge network data sharing," in *2022 IEEE International Conference on Unmanned Systems (ICUS)*, 2022, pp. 606–611.

[9] D. Mitropoulos, G. Gousios, P. Papadopoulos, V. Karakoidas, P. Louridas, and D. Spinellis, "The vulnerability dataset of a large software ecosystem," in *2014 Third International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, 2014, pp. 69–74.

[10] M. A. Dini, S. O. Ajakwe, I. I. Saviour, V. U. Ihekoronye, O. U. Nwankwo, I. U. Uchechi, G. A. Haryadi, M. A. P. Putra, D.-S. Kim, T. Jun *et al.*, "Patient-centric blockchain framework for secured medical record fidelity and authorization," , pp. 300–301, 2023.