2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE) | 979-8-3503-4501-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/CHASE60773.2024.00028

Decentralized Electronic Health Records Management via Redactable Blockchain and Revocable IPFS

Hao Guo* Wanxin Li† Collin Meese‡ Mark Nejad‡
*Research & Development Institute of Northwestern Polytechnical University in Shenzhen, China.
†School of Advanced Technology, Xi'an Jiaotong-Liverpool University, China
‡ Department of Civil and Environmental Engineering, University of Delaware, U.S.A.
haoguo@nwpu.edu.cn, wanxin.li@xjtlu.edu.cn, {cmeese,nejad}@udel.edu

***Abstract*—The increasing integration of the Internet of Health Things (IoHT) in the eHealth sector has significantly advanced the collection of Electronic Health Records (EHR). While blockchain technology offers enhanced data integrity and traceability for EHR, its inherent immutability often clashes with user concerns regarding security and privacy and prevents records from being securely updated. To address these challenges, our paper presents an innovative redactable blockchain mechanism tailored for EHR management. This approach leverages a decentralized, attribute-based chameleon hash function, enabling transaction-level redactions without succumbing to the vulnerabilities of a single point of failure. We implemented the proposed system based on the Charm cryptographic library and revocable IPFS scheme. The Chameleon hash function can support redactable operation for on-chain EHR-related information in seconds level. We also conducted extensive experiments on blockchain performance and IPFS performance, highlighting that the proposed redactable blockchain is efficient in practice for EHR management.**

I. Introduction

In recent years, the volume of electronic health data has increased dramatically, benefiting from advances in the development of smart health devices and the internet-of-health-things (IoHT). More specifically, IoHT devices are transforming how healthcare data is generated, stored, and used, resulting in significant research focused on improving security and ensuring privacy and accessibility for electronic health records (EHRs). EHRs have become vital for healthcare professionals, providing a database and data structure for storing patient-related medical information. However, with the advent of numerous cyber-attacks on healthcare data, a tremendous burden has been placed on healthcare providers and stakeholders to ensure the security and privacy of EHR data to protect their patients and comply with regulations, including the Health Insurance Portability and Accountability Act (HIPAA). For example, the HIPAA Journal reports that 5,510 healthcare data breaches exceeding 500 EHRs were reported to the U.S. Department of Health and Human Services Office for Civil Rights between 2009 and 2022, equating to the exposure of 382,262,109 healthcare records [1]. Furthermore, these privacy and regulatory concerns present challenges with data ownership and accessibility, where the safety mechanisms imposed by healthcare providers hinder a patient's ability to own and securely share their data amongst different healthcare networks. Consequently, there is a critical need for further research into developing improved methods for securing the storage and accessibility of EHR data.

Over the past decade, researchers have proposed numerous methods for improved EHR management systems. Despite considerable progress and due to the differences in legacy EHR systems across organizations, a superior and standardized system has yet to emerge. In the current landscape, healthcare networks utilize various methods for storing and managing EHR data, presenting challenges for patients who want to migrate their digital data to another network or grant access to out-of-network professionals. Most recently, blockchain-based methods have been explored as a secure, decentralized, and stakeholder-centric approach for improving accessibility and alleviating the burden of data management from healthcare organizations [2], [3]. Traditionally, blockchain networks are decentralized and immutable content storage and computing networks secured and managed by decentralized nodes [4]. They can be permissioned, where a consortium of stakeholders implements data access and membership controls, or permissionless, as in Bitcoin, where membership is open, and all participants can access data. Given its properties of programmable data access permissions and controls, permissioned blockchain technology presents a suitable candidate to improve the management and storage of EHR data.

In addition, blockchain technology holds the promise of revolutionizing traditional EHR management by providing a secure and privacy-preserving platform for storing and sharing sensitive medical information [5]. Unlike traditional centralized systems, where a single entity controls the data, blockchain networks distribute data across a network of nodes, making it extremely difficult for unauthorized parties to manipulate or access the information. The cryptographic principles underlying the blockchain ensure data integrity and confidentiality, further enhancing the security of EHRs.

Two key challenges exist despite the benefits of implementing permissioned blockchain technology in the EHR setting. The first challenge relates to the various formats of EHR information; it can contain textual, numeric, image, or video data. Image and video data are expensive to store directly

2832-2975/24/$31.00 ©2024 IEEE
DOI 10.1109/CHASE60773.2024.00028

on a blockchain network, and methods will be needed to store the raw data off-chain while securing it through linkage with the blockchain network. The immutability and tamper-resistant nature of the blockchain poses the second challenge, making updating historical EHR records or off-chain data links challenging. For example, when errors such as medical device malfunctions or human recording errors occur, it is necessary to be able to correct mistakes and update records and necessitate research into new methods for overcoming the challenges of realizing a blockchain-based EHR system.

To address the challenges in existing blockchain-based EHR management, this paper proposes a redactable blockchain mechanism for EHR management that includes a revocable InterPlanetary File System (IPFS) scheme for secure off-chain storage. Redactable blockchain methods leverage cryptographic techniques to enable controlled changes to blockchain records. Authorized users, such as healthcare providers or patients, can apply digital signatures and encryption to redact or update specific portions of the EHR or off-chain data links without compromising the security and immutability of the entire blockchain. This approach ensures compliance with regulations while maintaining the benefits of blockchain technology for EHR management and alleviating the burden of data management from healthcare organizations.

The remainder of the paper is structured as follows: Section II introduces the related work into blockchain-based EHR management systems; Section III presents the system architecture; In Section IV, we present the experimental results and evaluation of the system; and lastly, we conclude the study in Section V.

II. Related Work

Recently, researchers have investigated redactable blockchain applications for data sharing and management tasks. Zhang et al. [6] proposed a hierarchical access control redactable blockchain model for data sharing through attribute-based encryption, and the chameleon hash functions. Under the proposed model, the data owner can specify who can modify their data by setting an access policy and authenticating the modifier with a digital signature. However, they do not consider the off-chain storage and sensitive data revocation issues. Yeh et al. [7] proposed one comprehensive scheme integrating the redactable blockchain with the existing revocable IPFS mechanism. They have developed an enhanced proxy re-encryption scheme that simplifies access control for physicians without the need for complex group key management. However, their scheme required the centralized oracle to distribute public and private key pairs.

Li et al. [8] proposed a privacy-preserving dynamic searchable encryption framework named DSE-RB, a general scheme that guarantees reliable queries and updates on encrypted data. In particular, they also use the transaction-level editing mechanism to achieve a more flexible update operation of encrypted data without additional transactions while avoiding the waste of storage on the blockchain. However, they utilize the symmetric encryption algorithm (such as AES) to communicate between the data user and the data owner.

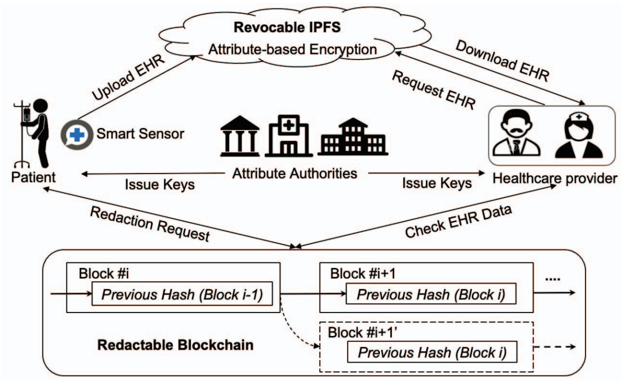

Figure 1: EHR Management System Architecture.

Jia et al. [9] proposed a decentralized and traceable redactable blockchain, which encodes redacted blocks in an RSA accumulator. Li et al. [10] proposed a redactable blockchain protocol that can be applied to proof-of-work and proof-of-stake blockchains and instantly redacted. Luo et al. [11] proposed ScalaCert to address the scalability issues of blockchain-based PKIs using a redactable blockchain to store the revocation information of the original certificate. Zhou et al. [12] proposed the redactable blockchain, which utilizes trapdoor-hash for transaction editing, deletion, and smart contract repair. They utilized holomorphic secret-sharing schemes to distribute the private key. Meanwhile, their schemes require heavy cryptographic computation and trusted authority.

To address the above issues, we propose a redactable blockchain architecture for EHR management supporting transaction-level redaction operations and also utilize the revocable IPFS scheme to protect the privacy of data owners.

III. System Architecture

This section presents the hybrid blockchain-IPFS architecture composing revocable IPFS and redactable blockchain mechanisms to secure patients' EHRs and protect private sensitive information. As shown in Fig. 1, we first describe all listed entities.

- EHR data: EHR data is the information owned by the patient and could be accessed by authorized and qualified healthcare providers who satisfy access policies. Also, these sensitive data could be removed on behalf of the patient.
- Patient: A patient is the data owner of their EHR and personal information; A patient can define access policies for data users (healthcare providers).
- Healthcare provider: The healthcare provider is a data user who wants to access EHR information. The healthcare provider actively requests access permissions from the patients (data owner).

- Attribute: An attribute is a piece of information (e.g., patient's ID) attested to the participants (data owner or the data user).
- Attribute authority: Multiple attribute authorities are entities that manage attributes and generate public/private keys to participants (data owners and users), which can later conduct the blockchain redaction operations.
- Smart sensor: A smart sensor is a device that collects and generates EHRs from data owners (patients) and then sends it to the IPFS storage, which is semi-trusted.
- Revocable IPFS: Revocable IPFS is the storage and computing platform that saves EHR information and other personal private data encrypted with the ABE scheme in a remote, semi-trusted place. It uses a Distributed Hash Table (DHT) and BitSwap technology for fast data storage and block distribution.
- Redactable blockchain: Smart contracts input patients' signatures and private keys and return the one-time EHR addresses if the access control policy has been satisfied. The redactable blockchain system records the EHR-related activities for data sharing and legitimate modification events.

In the remainder of this section, we first describe the decentralized attribute-based chameleon hash operation. EHR information presented in the paper is encrypted with ABE schemes, whereas the multi-authority authenticates the patients' private attributes (e.g., driver's license) anonymously and effectively. The proposed scheme is sufficient for the patient to prove their possession of attributes without revealing sensitive private information. Additionally, we propose on-chain and off-chain redaction architecture with a revocable IPFS platform. We also present detailed construction for redactable blockchain and revocable IPFS mechanisms.

A. Decentralized Attribute-based Chameleon Hash Operation

To address the centralized authority issue in traditional chameleon hash functions, we proposed the decentralized attribute-based chameleon hash (DACH) scheme. The DACH scheme is described as follows.

Syntax: We utilize the $Out \leftarrow Algorithm\ in$ to represent the protocol executed by p participants $P_1, ..., P_t$, where the input and output of P_i are in_i and out_i. The decentralized attribute-based chameleon hash function $DACH$ = $(AuthoritySetup, AuthorityKeyGeneration,$ $DACHKeyGen, Hash, Collision, Verify)$ includes the following seven protocols and algorithms.

Authority Setup$(IP, AT_i) \longrightarrow PK_i,\ SK_i$. Authority i takes the initial parameter IP and attributes AT_i as input value to generate public key pairs PK_i and secret master key SK_i.

Authority Key Generation$(GID, AT_i, SK_i, IP) \longrightarrow SK_i,_{GID}$. It takes as inputs attribute AT_i, global identification number GID of the participant p, secret master key SK_i, and the initial parameter IP, and finally outputs attribute key $SK_i,_{GID}$ to the participants p.

$pk_i, sk_i \leftarrow$ **DACHKeyGen**$(1^\lambda, 1^p)$. In this key generation phase, each participant P_i inputs the security parameter λ and the number of participants p. This algorithm outputs the attribute-based public key pk_i and private key sk_i share.

$(r, h) \longleftarrow$ **Hash**(pk_i, m). This hashing algorithm takes as input a public key pk_i and the message m. It will output the hash h and random value r.

$\{r'\}_{i=1}^p \longleftarrow$ **Collision**$(\{sk_i, pk_i, (m, r), m'\}_{i=1}^p)$. In this redaction algorithm, the inputs are each participant (attribute-based) its share private key sk_i, public key pk_i, random message (m, r), and redacted message m', it outputs the redacted random number r'.

$((r, h))$ = **Verify**$(pk_i, (m, r), (m', r'))$. The verification algorithm takes the input as public key pk_i, the original random message (m, r), the redaction message (m', r'), and it outputs the binary value which indicates the redaction message (m', r') is valid or not.

With the input of the public key g^p, the original random message $(m, r = (g^e, g^{pe}))$, and the redacted message $(m', r' = (g^{e'}, g^{pe'}))$, the verification algorithm will compute $d \longleftarrow Hash(g^p, m)$. Finally, it outputs 1 if $g^e d^m = g^{e'} d^{m'}$ and that $(g, g^p, g^{e'}, g^{pe'})$ is the Diffle-Hellman tuple and 0 otherwise.

Finally, the correctness of the decentralized attribute-based chameleon hash scheme requires that $pk_i, sk_i \leftarrow DACHKeyGen(1^\lambda, 1^p)$, all $(r, h) \longleftarrow Hash(pk_i, m)$, and that $\{r'\}_{i=1}^p \longleftarrow Collision(\{sk_i, pk_i, (m, r), m'\}_{i=1}^p)$. We have that $Hash(pk_i, (m, r)) = h$ and that $Verify(pk_i, (m, r), (m', r')) = 1$.

B. Redactable Operation

1) Merkle Tree: A Merkle tree is the data structure utilized in blockchain technology to organize and verify transactions within a block. Using hash functions like MD5 [13], or SHA-256 [14], each transaction is individually hashed to create a unique fingerprint. These hashed transactions are then combined in pairs and hashed again, continuing this process until a single hash, known as the Merkle root, is obtained. The Merkle root is stored in the block header, serving as an efficient representation and validation of the integrity of all the transactions within the block.

2) Transaction-level Redaction: Our blockchain implementation introduces a unique feature that supports redactions at the transaction level. We employ a modified version of the Merkle tree structure that utilizes the decentralized chameleon hash function to enable this functionality. Unlike traditional hash functions, decentralized chameleon hash functions possess the remarkable property of allowing transaction modifications without altering the corresponding hash values.

Utilizing the decentralized chameleon hash function, our blockchain enables the redaction of specific transactions within a block without necessitating changes to the associated hash values. This innovative approach allows for transaction editing while preserving the original hash values, ensuring the integrity and consistency of the Merkle tree structure. As shown in Algo. 1, Lines 1-3 present the preparation for redacting the transaction. Lines 4-8 present the redaction operation if it

```
Algorithm 1: Transaction Level Redaction Operation
  Input  : Transaction $t_i$ has been received for the
           redaction operation
  Output: New redacted transaction $t_i'$
1 while There exists a new transaction redaction request
    has been received for the blockchain do
2     Execute transaction redaction operation;
3 end
4 if $hash(t_i) == hash(t_i')$ then
5     redact the new transaction;
6 else
7     wait for another round of redaction operation;
8 end
9 The system updates the transaction $t_i$ into new
    redacted transaction $t_i'$;
```

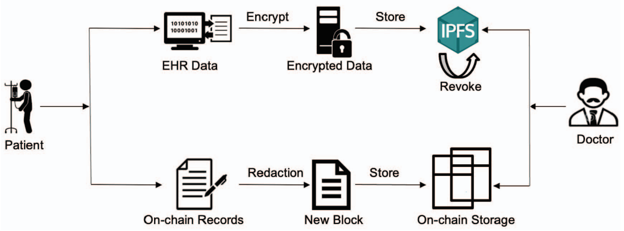

Figure 2: On-chain redaction records and off-chain revocable IPFS storage.

satisfies the predefined condition $hash(t_i) == hash(t_i')$. Line 9 outputs the new transaction t_i' while preserving the original hash values.

C. On-chain Redactable Records and Off-chain Revocable IPFS

As shown in Fig. 2, to address the security and privacy concern of personal EHR data [15], and the drawback of lacking revocation mechanism of IPFS. We develop a hybrid blockchain-IPFS architecture that includes the on-chain mechanism of block redaction operation. Also, the off-chain IPFS storage saves the ABE-encrypted EHR data, which can be revoked via legitimate operation. Blockchain transactions are transparent to each participant, saving the EHR redaction information and access activities.

Additionally, every patient can store personal information, including name, GID. The patient can pre-define the access policy for data users (e.g., healthcare providers) to access the EHR data, which is implemented with the access control policies of the blockchain system.

We propose the *RevokeData* function to revoke the existing EHR data from the IPFS. The revocable IPFS process first calculates the redaction message (m', r') by the $Collision()$ algorithm, which returns the redacted message m' and the random number r'. Next, the newly redacted message m' is stored in the blockchain transactions. Finally, one $UpdateTransaction$ request is sent to the redactable blockchain to erase the EHR information stored on the IPFS.

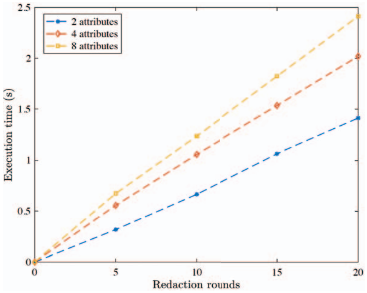

Figure 3: Execution time for DACH generation process.

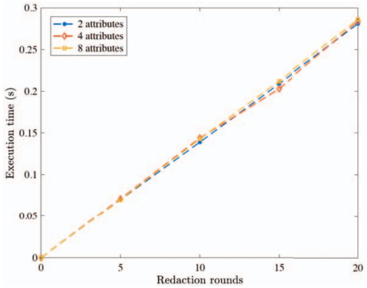

Figure 4: Execution time for DACH verification process.

IV. Experiments and Evaluations

A. Redactable Blockchain

The Chameleon hash function was prototyped in Python with the Charm cryptographic library [16]. We conducted a comprehensive performance evaluation of the proposed Decentralized Attribute-based Chameleon Hash (DACH). Specifically, our investigation focused on understanding how redaction rounds impact the execution time of DACH algorithms. We also compared these results across different numbers of DACH attributes, including 2, 4, and 8.

1) DACH Generation: Fig. 3 depicts the DACH hash collision generation time result. The DACH generation time will increase linearly when we increase the redaction rounds. Also, the DACH generation time will grow with the increasing number of attributes. The longest execution is around 2.5 seconds, with 20 rounds of redaction operation and 4 attribute conditions.

2) DACH Verification: Fig. 4 illustrates hash-verification time for various redactable operations and attribute counts. We found that all four lines are closed, and the average time cost for each redactable operation is nearly 0.14s. It indicates that incorporating decentralized attributes has less effect on the verification time. The reason is that the verification algorithm only checks the input hashed message, which is a fixed length.

B. Revocable IPFS

We evaluate the IPFS system's average write latency and read latency for the performance test. Network latency is

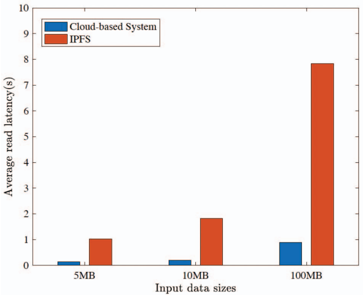

Figure 5: Comparison of read latency between cloud-based system and IPFS.

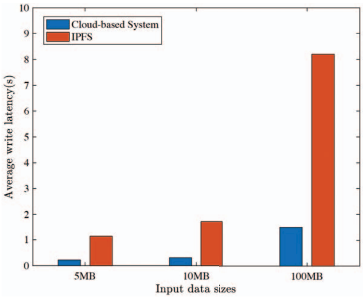

Figure 6: Comparison of write latency between cloud-based system and IPFS.

calculated by the time duration from invoking a request to receive the confirmation in the blockchain network. All experiments were conducted on the Ubuntu system with a 3.0 GHz Intel i7 processor and 16GB of memory.

1) Read Latency: We compare the average read latency with cloud-based systems and IPFS with different input data sizes (5MB, 10MB, and 100MB). As we can see from the Figure. 5, cloud-based systems take less time when conducting reading tasks. When the data size is 10MB, IPFS takes 1.8s, while cloud-based systems take 0.2s.

2) Write Latency: We also compare the average write latency with cloud-based systems and IPFS with different input data sizes (5MB, 10MB, and 100MB). As we can see from the Figure. 6, the cloud-based systems also take less time when conducting the writing task. When the data size is 10MB, IPFS takes 1.7s, while cloud-based systems take 0.3s. From the experiment observation, the IPFS read/write execution times are acceptable when increasing the input data sizes (all less than 10s).

V. Conclusion

This paper proposed a redactable blockchain solution for EHR data management. The proposed approach utilizes a decentralized and attribute-based chameleon hash scheme to provide secure modification of on-chain EHR-related information. In addition, the revocable IPFS scheme enables efficient storage and secure modification of off-chain EHR data that cannot be efficiently stored on-chain. We conducted extensive experiments to verify the performance of our proposed architecture based on the Charm cryptographic library and revocable IPFS scheme. The results indicate that the redaction operations can be processed in seconds level. At the same time, the IPFS read and write latency is on the order of seconds and varies based on the input data.

Acknowledgment

This work is partially supported by the Guangdong Basic and Applied Basic Research Foundation under the Grant No. 2021A1515110286, 2021-2024.

References

[1] R. Murray-Watson, “Healthcare data breach statistics,” 2023. [Online]. Available: https://www.hipaajournal.com/healthcare-data-breach-statistics/

[2] H. Guo, W. Li, M. Nejad, and C. Shen, “Access control for electronic health records with hybrid blockchain-edge architecture,” in *2019 IEEE International Conference on Blockchain (Blockchain)*, 2019, pp. 44–51.

[3] B. Zaabar, O. Cheikhrouhou, F. Jamil, M. Ammi, and M. Abid, “Healthblock: A secure blockchain-based healthcare data management system,” *Computer Networks*, vol. 200, p. 108500, 2021.

[4] H. Guo, C. Meese, W. Li, C.-C. Shen, and M. Nejad, “B2sfl: A bi-level blockchained architecture for secure federated learning-based traffic prediction,” *IEEE Transactions on Services Computing*, pp. 1–15, 2023.

[5] W. Li, C. Meese, H. Guo, and M. Nejad, “Aggregated zero-knowledge proof and blockchain-empowered authentication for autonomous truck platooning,” *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[6] T. Zhang, L. Zhang, Q. Wu, Y. Mu, and F. Rezaeibagha, “Redactable blockchain-enabled hierarchical access control framework for data sharing in electronic medical records,” *IEEE Systems Journal*, vol. 17, no. 2, pp. 1962–1973, 2023.

[7] L.-Y. Yeh, W.-H. Hsu, and C.-Y. Shen, “Gdpr-compliant personal health record sharing mechanism with redactable blockchain and revocable ipfs,” *IEEE Transactions on Dependable and Secure Computing*, pp. 1–16, 2023.

[8] M. Li, C. Jia, R. Du, W. Shao, and G. Ha, “Dse-rb: A privacy-preserving dynamic searchable encryption framework on redactable blockchain,” *IEEE Transactions on Cloud Computing*, vol. 11, no. 3, pp. 2856–2872, 2023.

[9] M. Jia, J. Chen, K. He, R. Du, L. Zheng, M. Lai, D. Wang, and F. Liu, “Redactable blockchain from decentralized chameleon hash functions,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2771–2783, 2022.

[10] X.-Y. Li, J. Xu, L.-Y. Yin, Y. Lu, Q. Tang, and Z.-F. Zhang, “Escaping from consensus: Instantly redactable blockchain protocols in permissionless setting,” *IEEE Transactions on Dependable and Secure Computing*, pp. 1–20, 2022.

[11] X. Luo, Z. Xu, K. Xue, Q. Jiang, R. Li, and D. Wei, “Scalacert: Scalability-oriented pki with redactable consortium blockchain enabled” on-cert” certificate revocation,” in *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2022, pp. 1236–1246.

[12] G. Zhou, X. Ding, H. Han, and A. Zhu, “Fine-grained redactable blockchain using trapdoor-hash,” *IEEE Internet of Things Journal*, 2023.

[13] Z. Yong-Xia and Z. Ge, “Md5 research,” in *2010 second international conference on multimedia and information technology*, vol. 2. IEEE, 2010, pp. 271–273.

[14] H. Gilbert and H. Handschuh, “Security analysis of sha-256 and sisters,” in *International workshop on selected areas in cryptography*. Springer, 2003, pp. 175–193.

[15] H. Guo, W. Li, M. Nejad, and C.-C. Shen, “A hybrid blockchain-edge architecture for electronic health record management with attribute-based cryptographic mechanisms,” *IEEE Transactions on Network and Service Management*, pp. 1–1, 2022.

[16] J. A. Akinyele, C. Garman, I. Miers, M. W. Pagano, M. Rushanan, M. Green, and A. D. Rubin, “Charm: a framework for rapidly prototyping cryptosystems,” *Journal of Cryptographic Engineering*, vol. 3, pp. 111–128, 2013.