CHAPTER 1

INTRODUCTION

The implementation of Know Your Customer (KYC) processes is essential for banks and financial institutions to verify the identities of their customers [5]. With technological advancements, many institutions are transitioning from traditional KYC to electronic KYC (e-KYC) systems. However, this shift has introduced new challenges related to efficiency, effectiveness, and customer experience [4]. Customers often find the data entry process cumbersome, and institutions face difficulties in managing and securing the large volumes of sensitive data involved.

Furthermore, the use of cloud-based e-KYC systems raises significant privacy concerns. Storing sensitive customer data in the cloud increases the risk of unauthorized access and data breaches [11]. While some banks implement encryption and decryption mechanisms on the cloud side to enhance security, this introduces issues related to key distribution and centralized validation, particularly concerning key revocation [4].

As the volume of data continues to grow, efficient verification becomes increasingly critical. Traditional methods, which require files to be verified individually, are both time-consuming and inefficient. Batch verification could offer a solution by allowing multiple files to be verified at once, but existing e-KYC systems lack effective batch verification capabilities, particularly in decentralized environments such as blockchain networks [4].

This study addresses these challenges by proposing the use of searchable symmetric encryption (SSE)—a cryptographic method that allows encrypted data to be searched efficiently—integrated with blockchain technology. To further enhance the security and efficiency of this approach, the research introduces a novel enhancement to the PRNG within the SSE framework by employing turbulence-padded chaotic maps. This enhancement leverages the unpredictability of chaotic systems to strengthen the cryptographic properties of the SSE, improving verification speed and maintain privacy protection within e-KYC systems. Ultimately, this approach aims to maintain privacy, security, and improving operational efficiency, thus improving both the customer experience and institutional processes.

1.1 Rationale

The financial sector universally acknowledges the critical role of the Know Your Customer (KYC) process in verifying client identities and ensuring compliance with regulatory standards [5]. With the rapid advancement of digital technologies, there has been a global shift from traditional paper-based KYC procedures to electronic KYC (e-KYC) systems. This transformation aims to enhance efficiency, reduce operational costs, and improve the overall customer experience. However, this shift has introduced new challenges that are being felt globally, nationally, and locally.

Globally, financial institutions are grappling with inefficiencies and security concerns associated with e-KYC processes [4]. Customers often find the data entry phase during e-KYC registration cumbersome and time-consuming, leading to dissatisfaction and potential attrition. Nationally, the proliferation of disparate e-KYC systems developed independently by various banks and financial institutions has resulted in a lack of standardization, interoperability issues, and increased complexity in data management.

Governments use e-KYC systems to verify the authenticity of citizen documents such as national IDs, passports, and financial records. The validation process involves multiple layers of security, automated checks, and integration with national databases to ensure accuracy and prevent fraud.

Locally, institutions deploying e-KYC systems on cloud platforms face significant privacy and security risks. Storing sensitive customer data in the cloud elevates the potential for unauthorized access and data breaches [4]. While encryption mechanisms are implemented to safeguard data, they introduce complexities related to key distribution, management, and revocation, especially when relying on centralized validation systems [4].

Authoritative sources stress the urgent need for more robust, efficient, and secure systems to handle the escalating volume of KYC data and associated risks. For instance, [3]. highlight the effectiveness of Searchable Symmetric Encryption (SSE) in enabling secure search operations over encrypted data without compromising privacy [3]. Additionally, Liu et al. discuss the challenges of data retrieval in decentralized storage systems and the importance of efficient search mechanisms [10].

In light of these challenges, there is a pressing need for innovative solutions that enhance both the efficiency and security of e-KYC processes. This research addresses this need by proposing the integration of Searchable Symmetric Encryption (SSE) with blockchain technology to enable efficient batch verification of e-KYC documents. By conducting experiments using a simulated blockchain network on a local computer and employing a news corpus to mimic large datasets, this study aims to demonstrate the capability of SSE to handle big data scenarios effectively.

This approach not only seeks to mitigate existing inefficiencies and security concerns but also aims to improve e-KYC processes globally. By enhancing operational effectiveness and customer satisfaction, the proposed solution could significantly impact the financial sector, offering a scalable and secure method for handling the ever-growing demands of KYC compliance.

1.2 Theoretical Framework

The theoretical framework of this study is based on the principles of searchable symmetric encryption (SSE) and its application within blockchain networks for secure data management and retrieval. SSE is a cryptographic scheme that allows users to perform secure searches over encrypted data without revealing the actual data to unauthorized parties [3]. This capability is particularly useful in scenarios where sensitive data is stored in untrusted environments, such as cloud storage or decentralized networks.

In the context of e-KYC systems, SSE facilitates the secure batch verification of encrypted documents. By encrypting and indexing data, SSE enables efficient search and retrieval operations while maintaining data confidentiality. This is achieved by generating pseudo-random keys through stream ciphers, which are used to encrypt and extract keywords or indexes from the original files [3].

Blockchain technology provides a decentralized and tamper-resistant platform for storing and managing data. Integrating SSE with blockchain enhances data security and integrity, as the immutable ledger ensures that data cannot be altered without detection. This combination addresses the privacy concerns associated with cloud-based e-KYC systems and improves the efficiency of verification processes.

1.3 Conceptual Framework/Paradigm

The key variables related to the problem of batch verification of e-KYC documents in blockchain networks include the e-KYC system, encryption techniques, privacy concerns, and batch verification. The e-KYC system is a crucial variable, as it forms the basis of the customer identification and verification processes. Encryption techniques, such as searchable symmetric encryption, directly impact the security and privacy of customer data in the e-KYC process. Privacy concerns are another critical variable, as they underline the need for improved data protection and user traceability. Lastly, the batch verification process is a fundamental variable as it serves as the focus of the research problem, aiming to improve the efficiency of verifying multiple e-KYC documents simultaneously.

A schematic diagram of the paradigm of the research can be represented as follows:

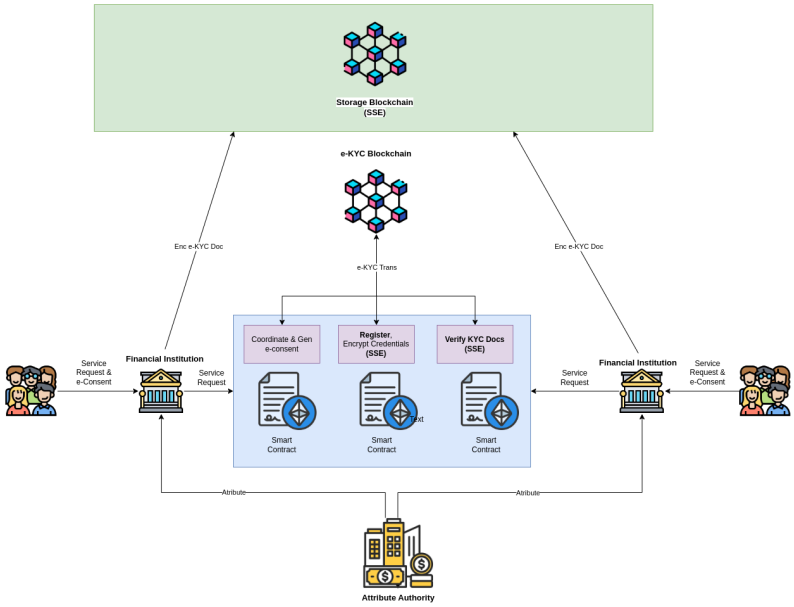

Figure 1.1: SSE General Schema

In this diagram, the e-KYC system is built on the basis of 2 blockchains that hold each specific data. Decentralize storage data for encrypted Doc and e-KYC Blockchain data transaction and validation. Encryption techniques, including searchable symmetric encryption, are integrated with the e-KYC system, emphasizing their role in safeguarding sensitive information. Privacy concerns are linked to encryption techniques, underscoring the importance of addressing data privacy issues through effective encryption methods. Finally, the batch verification process is interconnected with the e-KYC system and encryption techniques, illustrating its reliance on these variables to streamline the verification of multiple documents.

This paradigm demonstrates the intricate relationship and interdependence of the variables related to the problem, ultimately highlighting the holistic approach required to address the challenges in batch verification of e-KYC documents in blockchain networks.

1.4 Statement of the Problem

Batch verification remains a significant challenge in existing e-KYC systems, as highlighted by Fugkeaw [4]. Current methods require files to be verified individually, which limits efficiency and becomes increasingly cumbersome as the number of documents grows. An effective batch verification process would allow multiple files to be verified simultaneously, streamlining operations and reducing processing time. Furthermore, as the volume of stored files within the Decentralized Hash Table (DHT) continues to expand, data retrieval becomes increasingly complex and time consuming, further impeding the verification process [10]. This research aims to address these issues by developing an optimized solution for batch verification in e-KYC systems,using SSE because it fast and secure[13].

1.5 Objective and Hypotheses

One of the primary goals of the proposed method is to enhance batch verification performance significantly. Instead of carrying out only one verification per request, the method aims to process multiple verifications within a single request. This shift from a one-to-one verification model to a one-to-many paradigm not only increases throughput, but also reduces network overhead and latency. As a result, large-scale systems can benefit from faster and more resource-efficient verification processes, ultimately enabling real-time or near real-time validation for a high volume of data or transactions.

SSE is employed to search for encrypted fileID hash data on a Distributed Hash Table (DHT), both for single and multiple files [10]. In SSE, the search for an encrypted file is conducted by encrypting and extracting words or indexes from the original file. A pseudo-random number generator enables the creation of a key schema for the stream cipher used in SSE [3]. This approach facilitates file searches on the blockchain and allows users to proofread files they wish to decrypt [3].

The core method in this research is SSE, with a particular emphasis on improving its standard stream cipher through a modification of the pseudo-random number generator using Turbulence-Padded Chaotic Maps (TPCM). This modification aims to further optimize SSE's performance for batch verification, resulting in more efficient, secure, and scalable document handling in blockchain-based e-KYC environments. By exploiting the inherent properties of TPCM—especially its high sensitivity to initial conditions and its unpredictability—the algorithm can maintain robust security. These chaotic characteristics expand the key space and make brute-force or pattern-based attacks significantly more difficult, thereby reinforcing the confidentiality and integrity of the data.

1.6 Assumption

The underlying assumption of this research is that integrating batch verification through Searchable Symmetric Encryption (SSE) in e-KYC procedures will address existing inefficiencies associated with verification processes. By enabling the simultaneous verification of multiple documents, SSE is expected to streamline operational workflows and expedite the overall verification pipeline.

Additionally, it is assumed that using a dataset of 1,000 synthetic KYC records will effectively emulate real-world e-KYC environments that handle large amounts of customer data. These records are designed to mirror typical KYC documents in terms of complexity,

file sizes, and diversity of data fields. Employing synthetic data allows for a thorough assessment of SSE's capability to handle substantial workloads without compromising the privacy of real customers' information or violating regulatory standards.

By leveraging these 1,000 synthetic KYC documents, the study posits that any observed performance metrics, challenges, or enhancements during the batch verification process will closely resemble those encountered with genuine KYC data. This setup enables a realistic evaluation of SSE's scalability and robustness, providing insights into how the technology might enhance verification efficiency in actual financial institutions. Ultimately, the research assumes that SSE, tested on a significant synthetic dataset, will demonstrate meaningful improvements in verification speed, security, and data handling—reinforcing its potential suitability for widespread implementation in e-KYC systems.

By implementing this SSE method is just replace the IPFS blockchain with blockchain that support with SSE to handle encryption data and all the feature of trustblock by fugkeaw is still not change, including the CP-ABE privacy preserving mechanism.

For the guessing attack, it is assumed that at least 8 tokens are used, providing a security level comparable to a 256-bit hash. If fewer tokens are used, the security margin decreases, making attacks more feasible. Similarly, increasing the number of tokens—such as using 10 tokens to achieve 2^{320} security—significantly strengthens protection against brute-force attacks.

Governments use e-KYC systems to verify the authenticity of citizen documents such as national IDs, passports, and financial records. The validation process involves multiple layers of security, automated checks, and integration with national databases to ensure accuracy and prevent fraud, the user still need to upload the cred files to make it as a consent.

1.7 Scope and Delimitation

The scope of this experiment involves evaluating the efficiency of Searchable Symmetric Encryption (SSE) in handling large datasets of Know Your Customer (KYC) documents. Specifically, this study simulates a dataset consisting of 1,000 KYC files, using a news corpus to mimic a large volume of real KYC data. The primary objective is to assess the effectiveness of SSE in processing and searching through large KYC datasets, focusing on the speed and accuracy of searches in batch verification processes. The experiment is conducted within a controlled environment using a local computer network, simulating a blockchain network. This approach allows for a detailed observation of SSE's performance in handling large datasets and its impact on the batch verification process. By testing with a substantial dataset, the study aims to address the current challenges faced by electronic KYC (e-KYC) systems, particularly in financial institutions, where verifying multiple documents simultaneously can be time-consuming and resource-intensive.

The research aims to demonstrate how SSE can improve the efficiency of searching and verifying KYC documents, while ensuring data privacy and integrity. This setup is intended to simulate the real-world KYC search process and offer insights into the scalability of SSE for use in handling big data contexts, such as those found in financial services and other sectors that require robust identity verification processes. Delimitations of the study include the focus on simulated data, which, while providing a controlled test environment, may not fully capture all complexities of real-world KYC data. Additionally, the experiment is limited to the use of SSE in conjunction with blockchain technology and

does not include comparisons with other encryption techniques or decentralized storage systems outside the context of the blockchain network simulated here.

1.8 Significance of the Study

This study introduces a novel approach by modifying the pseudo-random number generator within the Searchable Symmetric Encryption (SSE) framework using a turbulence-padded chaotic map, then integrating SSE with blockchain technology to improve the secure searching process within the blockchain. By leveraging the high sensitivity and unpredictability of chaotic maps, the approach strengthens cryptographic security while significantly accelerating data retrieval in the blockchain system. This synergy not only optimizes the performance of e-KYC verification but also ensures that stringent security measures are upheld throughout the data handling and verification stages.