



# Deduplication Reduction in E-Voting system using Blockchain

K. Ashok Kumar<sup>1</sup>, K. Kasiviswanath<sup>2</sup>, P. Nithish Kumar Reddy<sup>3</sup>, V. Mahendra Kumar<sup>4</sup>

Assistant Professor<sup>1</sup> and Students<sup>2,3,4</sup>

Dhanalakshmi College of Engineering, Chennai, India

**Abstract:** *There have been several other voting methods since then. Paper ballots are the most widely used kind worldwide. Issues related with electronic voting methods have just become common in the last ten years. Electronic voting systems have issues primarily with security, reliability, trustworthiness, transparency, and functionality. The most advanced nation in this area is Estonia, a pioneer in the industry. There are, however, relatively few blockchain-based alternatives. All of those problems can be solved with blockchain, which also has certain benefits like immutability and decentralisation. Blockchain technologies for voting via electronic means have a number of issues, including a limited focus or a dearth of testing and benchmarking. This article introduces a blockchain-based electronic voting system that may be applied to any type of vote. It is totally driven by blockchain and capable of managing all procedures. The platform functions as entirely independent and decentralised once voting has begun, avoiding any potential problems with the voting process. The voter identities are protected by homomorphic encryption, but the data is completely visible. We examined and contrasted our solution across three distinct blockchains. According to the findings, there is minimal speed difference between using public and private blockchains. The main innovation of our system is the fully decentralised management of the electronic voting platform via blockchain, the transparency of the entire process, and simultaneously the protection and privacy of voters owing to homomorphic encryption.*

**Keywords:** Electronic voting, smart contracts, blockchain, ethereum, Hyperledger Composer, elections, homomorphic encryption.

## I. INTRODUCTION

The development of electronic voting systems is still in its infancy. We picked this topic because it is timely and since there aren't many solutions

available to the issues with streamlined voting. These days, e-Government development is becoming more and more popular. However, until fundamental citizen functions like elections are conducted electronically, such a system is not practical. One of the important public areas that blockchain technology has the potential to alter is electronic voting [1]. Electronic voting brings with it novel challenges that must be resolved. One of them, for instance, is for example election security, which must be as least as secure as traditional voting approaches using ballots. We made the conscious choice to design secure elections so that voters wouldn't have to worry about someone manipulating the democratic process. Blockchain has been cited consistently in recent years as an example of a highly encrypted technology utilised in the internet space. Our electronic voting system manages all voting procedures using blockchain. Its key benefit is that there is no requirement to believe in the centralised entity that organised the election. In our system, this authorization cannot change the outcome of an election. The lack of transparency in how the system operates is another issue with computerised voting. It causes voters to lack confidence [2]. Blockchain offers a transparent solution to this issue that makes it possible for everyone to view the methods used to store data and how the content is handled. This solution is far superior to the traditional e-voting platform without blockchain in terms of security.

The following is the breakdown of the article. A brief examination of current blockchain e-voting options is presented in Part II. In section III, we discuss everything and demonstrate the conceptual framework of our solution.

## II. LITERATURE SURVEY

The most important phase of the software development process is literature research. It's important to assess the company's strength, the economics, and the time factor before building a tool. The next stage is to choose the operating system and language that may be utilised to construct the tool when these have been finished. Programmers require a lot of outside support once they begin creating a product. You can find this assistance online, in books, or from veteran programmers. The previously mentioned variables should be taken into account prior to system development for the proposed system.

A large portion of the project development industry carefully evaluates and investigates each need needed for project development. The most crucial phase of the software development process for any project is literature research. Prior to the creation of the tool and the accompanying design, it is necessary to establish and investigate the time factor, resource needs, personnel, economics, and corporate strength. Once these criteria have been satisfied and thoroughly investigated, the following step is to establish the software requirements for the system in question, including what kind of operating system the project would need and what other software is required to move on to the following phase, which includes tool development and associated operations.

[1] In this research, they propose an architecture and then implement it in a system called dupLESS that offers a secure deduplication repository that is immune to brute force assaults. allows customers to use existing services with encrypted data. Performance and space savings from encryption for deduplicated storage can be comparable to those of a storage service using plain text data.

[2] A way exists to free up space from random duplication so that controlled file replication can use it. Even though the files are encrypted with different users' keys, the convergent encryption process allows duplicate files to be combined into the single file.

[3] In this basic strategy, the convergent keys are encrypted and sent to the cloud by each user using their own master key. However, if the number of users grows, such a simple key management approach creates a lot of keys, necessitating that users only safeguard their master keys.

[4] On the basis of accepted cryptographic presumptions, this project develops a private deduplication mechanism. The following presents and analyses it. They demonstrate that discrete-log coding techniques may erase up to a high proportion of bits and that protocols for re-deleting private data are likely to be secure, provided that the underlying hash function is collision-resistant. Here I am. [5] In this research, we create an encryption system that enhances storage and bandwidth advantages while ensuring semantic security for sensitive data while weakening the security of popular data. This means that although semantically secure encryption protects intangible material, data removal is effective for popular data. We demonstrate that, given the Diffie-Hellman symmetric outer decision assumption, our technique is safe.

## III. SOFTWARE IMPLEMENTATION

Python: Created by Guido Van Rossum in 1991, Python is an interpreted, object-oriented programming language. In contrast to other languages, it produces less code.

## IV. EXISTING SYSTEM

Private clouds are used as proxies in data deduplication systems, enabling data owners and users to safely carry out duplicate checks with different levels of access. These structures are useful and have drawn a lot of interest from scientists. While data operations are maintained in the private cloud, data owners simply outsource data storage utilising the public cloud.

A sophisticated data compression technique called data deduplication is used to get rid of redundant copies of frequently used data in storage. This method is used to increase storage efficiency and may also be used to send network data more efficiently by sending fewer bytes. Instead of maintaining numerous copies of the same data, deduplication eliminates redundant data by retaining only one physical copy and pointing supplementary redundant data to that copy.

Deduplication can be carried out at the block or file level. Duplicate copies of the same file are eliminated via file-level redo. The block level can also be used for deduplication. chunks of data that are duplicated in non-identical files are removed. various ciphertexts come from identical copies of data for various users, making copying impossible.

## V. DESIGN

The suggested blockchain voting system is intended to be used for all elections, including presidential, parliamentary, and student elections. It takes into consideration all voting needs. The system leverages public blockchain preferably and enables for multi-round elections. Other forms of blockchain can take the place of the public blockchain, but each user must be able to quickly verify the recorded data (votes). Any observer who wants to cast a vote on the blockchain is represented by a user.

We identify three crucial roles in our proposed system: voter, key authority, and publisher voice. These three positions can stand in for a group, a business, or a user. Because they may belong to the same entity or individual, the publisher and key authority voting responsibilities can be combined into a single position.. Depending on how the votes are arranged, the voter takes part in the election. The vote issuer configures the votes, which is done as part of the smart contract. Before publishing the smart contract, the voting issuer must be familiar with the encryption keys. The key authority and the poll issuer must work closely together. All encryption keys are generated and given to voters and voters by the key authority. The distribution route must be safe and impervious to outside interference.

Voting system's high-level architecture seen in Figure 1. Here are the functions, elements, and connections between them. The following elements are part of the architecture: the blockchain (required), the eID (required), the results interface (required), and the key authority API (optional). A smart contract that is a part of the blockchain and is in charge of processing and analysing votes is a specialised component. The kind of blockchain determines the smart contract's programming language.

A Key Authorization organisation may include an optional module called the Key Authorization API. The public key for homomorphic encryption and the key to access the votes are provided via this API to the poll publisher and voter, respectively. You can hand out these keys manually or in various ways. Because of this, the component in the architecture is simply suggested but not necessary.

The blockchain component executes voting procedures and represents the whole architecture for data storage. Blockchain may be created using either a private blockchain like Hyperledger or a public blockchain like Ethereum. A public blockchain has the benefit of making all information about transactions and blocks available to all users, which makes it more trustworthy than a private blockchain.

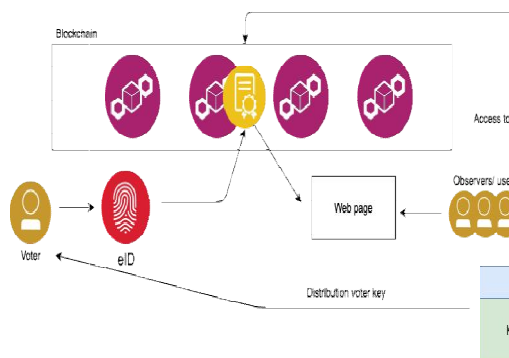
This trust is being used in a typical setting. The same degree of confidence may be provided by a private blockchain, but organisations must demonstrate it by providing the data. The application cases for blockchain are not constrained by the suggested architecture. The same level of trustworthiness may be offered by both kinds of blockchain. The blockchain organisation chooses the platform to be utilised.

A smart contract that is a component of the blockchain ensures the processing and bases the security of the vote on it. After setup, the smart contract is published on the blockchain network.

There are times, candidates, and other attributes in a configuration. A candidate can be anything that is the topic of an election; they don't even have to be actual people. Voting is more transparent since the published smart contract cannot be altered or updated. A list of users with voting rights is contained in the smart contract.

The access list needs to adhere to the key distribution policies put in place by the key authority.

The data is encrypted using homomorphic encryption and kept on the blockchain [10]. To encrypt the election results, the smart contract has to have a homomorphic encryption public key. The results are shown using the private key, which is kept separate and is utilised once voting is complete. This key can be given to several parties in charge of scrutinising votes. No one is aware of the findings since they are only apparent at the conclusion. We employed a zero-knowledge proof [11] to increase security by confirming that the votes include the right values, such as the restriction that a voter may only cast one vote and cannot cast two or more. For the purpose of voting, this component offers access to the blockchain. An eID that serves as a gateway to the blockchain network is taken into account in the architecture.



The key authority issues the user's public and private keys, which are contained on the ID card.

The standard public address of the wallet's blockchain wallet is represented by the public key. The private key is only known by the voter.

Results Interface (D) The results interface is represented by this component. The interface has to be able to access the blockchain and give users and observers information. Voting results are included in the record, and all transactions on the blockchain should be accessible for viewing. For ease of interpretation, the findings are presented below in graphical form, and just the final results are visible.

## VI. PROPOSED SYSTEM

The system is secured and hardened in the suggested work. We specifically offer a sophisticated method that offers increased security by encrypting data with keys that have various levels of privilege. Users without the necessary rights are prevented from doing repeated checks in this way. Furthermore, even when working along with the S-CSP, these malicious individuals are unable to decipher the ciphertext. According to the definitions given in the proposed security model, the security analysis demonstrates that the system is secure.

It is suggested to use convergent encryption to protect data privacy and provide redundancy. uses a converged key to encrypt or decode a copy of a data file after calculating its cryptographic hash value.

The user then maintains the key and uploads the encrypted data to the cloud after creating the key and encrypting the data. Identical copies of the data yield the same convergent key and, consequently, the same ciphertext since cryptographic procedures are predictable and derived from the data content. To A secure proof-of-ownership mechanism is also required to show that the user genuinely owns the same file if duplicates are discovered in order to prevent unauthorised access.

## VII. EVALUATION

After the design process, we made the decision to put the necessary service into place. We developed two test scenarios to ensure the functionality of our software. The primary objective was to evaluate the solution's general functioning, security, and speed. A MacBook Pro running macOS Mojave (10.14.3), the Chrome web browser with JavaScript enabled, and the MetaMask plugin were the components of our test setup. A straightforward interface for interacting with the Ethereum network is provided by MetaMask. We built a small Ethereum network with Ganache for testing reasons. We released the smart contract to the Ropstentestnet after all automated truffle testing were successful.

The first test case contains fundamental unit tests that verified the correctness of the smart contract's behaviour. 1000 simulated test users who typically voted like actual people were used to implement it. Additionally, tests were run.

The second test case was timing how long it would take a real individual to cast a ballot. The plan was to launch the voting web application, select the candidate you wanted to support, and then vote. The vote would then be added to the blockchain and registered. The smart contract was implemented in three separate blockchains for the test: Ganache, a local Ethereum network, Hyperledger Composer, a private blockchain, and Ropsten, a live Ethereum test network. A background script was automatically voting 15 persons each second.

**VIII. EXPERIMENTAL RESULT:**

This Section Demonstrates the performance of the project programme as a result of its implementation of this undertaking

```

In [5]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_style('white')

In [6]: pd.read_excel("C:\Users\UTSS\Downloads\csv_duplicate_removal-20230307060312-001\csv_duplicate_removal\datavert.xlsx")

In [8]: file_df.info()

Out[8]:
Out[8]: <class 'pandas.core.frame.DataFrame'>
Int64Index: 581 entries, 0 to 580
Data columns (total 6 columns):
 #   column      non-null count  dtype
---  ---
 0   name        581 non-null    object
 1   father/husband  581 non-null    object
 2   age         581 non-null    int64
 3   gender      581 non-null    object
 4   voterId     580 non-null    object
 5   hadhar no   581 non-null    object
dtypes: int64(1), object(5)
memory usage: 22.4k B

In [10]: file_df.isnull().sum()

```

```

In [35]: sms.set_title('data')

In [36]: file_df.first_record.to_excel("C:\Users\UTSS\Downloads\csv_duplicate_removal-20230307060312-001\csv_duplicate_removal\visu")

In [37]: file_df.last_record = file_df.drop_duplicates(subset=["name", "father/husband", "age", "gender", "voterId", "hadhar no"], keep="last")

In [38]: sms.pairplot(file_df, hue = "gender")
plt.show()

```

```

In [17]: file_df.last_record = file_df.drop_duplicates(subset=["name", "father/husband", "age", "gender", "voterId", "hadhar no"], keep="last")

In [18]: sms.pairplot(file_df, hue = "gender")
plt.show()

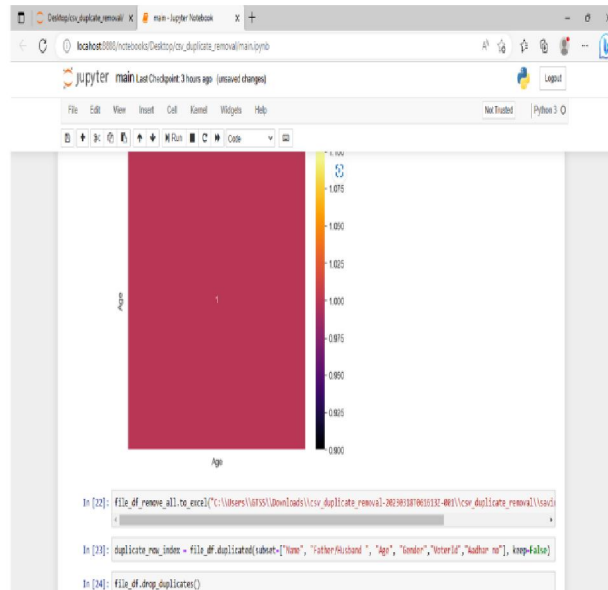
In [19]: file_df.last_record.to_excel("C:\Users\UTSS\Downloads\csv_duplicate_removal-20230307060312-001\csv_duplicate_removal\visu")

In [20]: file_df_remove_all = file_df.drop_duplicates(subset=["name", "father/husband", "age", "gender", "voterId", "hadhar no"], keep="all")

In [21]: sms.kstest(file_df_corr(), annct = True, cnsap = 'inferno')
plt.show()

```





	Name	Father/Husband	Age	Gender
0	Prasanna Lakshmi Chinnappa Reddy	Siva Reddy C	25	F
1	Gundeethi Mallikarjuna	Bhadrath	66	M
2	Padmaavathi	Chinnappa	41	F
3	Vidya Chinnappa Reddy	Rameshanna Reddy	30	M
4	Parvathamma Chinnappa Reddy	Venkata Siva Reddy	54	F
...	...	...	...	...
485	KANCHARLA MADHVI CHONDREY	KANCHARLA NAGESH BABU	24	M
486	KONDILLA FAJRESH BABU	KONDILLA VENKATA RAU	23	M
487	KONISETTY PAVANI KALYANI	KONISETTY VENKATESHVARULU	23	M
488	KURAVAM SIVA RAJESH	KURAVAM MURALI KRISHNA	22	M
489	KARNATI RAJAHANBURA REDDY	KARNATI ASHOK REDDY	22	M

	Name	Father/Husband	Age	Gender
0	Prasanna Lakshmi Chinnappa Reddy	Siva Reddy C	25	F
1	Gundeethi Mallikarjuna	Bhadrath	66	M
2	Padmaavathi	Chinnappa	41	F
3	Vidya Chinnappa Reddy	Rameshanna Reddy	30	M
4	Parvathamma Chinnappa Reddy	Venkata Siva Reddy	54	F
...	...	...	...	...
376	Krishna Reddy	Pulla Reddy	39	M
377	Varalakshmi	Srinivasa Rao	44	F
...	...	...	...	...

We ran this script in an effort to recreate the actual circumstances that may arise during actual elections. The test scenario was run five times for each blockchain, and Table I displays the test results. Be aware that the Ropsten network's average blocking time, which is about 12 seconds, has an impact on the difference in timings.

### IX. CONCLUSION

Although there are little variations in network delays, they are so minor that public blockchain offers greater advantages in this type of voting system since the data is accessible and anybody can view it in real time. A private blockchain is a little bit quicker, but because it only operates where the authority wants it to, it degrades the credibility of the entire system. According to the table, it takes Ganache 6.32 seconds (median 6.34 seconds), Hyperledger Composer 6.05 seconds (median 6.04 seconds), and Ethereum Ropsten 17.75 seconds (median 17.93 seconds) on average to add one vote. Both the block time and the chosen consensus algorithm have an impact on these periods.

### X. CONFIRMATION

The Incentives for Research and Development programme of the Ministry of Education, Science, Research and Sport of the Slovak Republic provided funding for this study under grant agreement 2018/14427:1-26C0. Additionally, it belongs to the APVV-15-0731 project. The financial support for young researchers provided by the STU Grant Scheme is appreciated by the authors

### REFERENCES

- [1]. N. Kshetri and J. Voas,
- [2]. "Blockchain-Enabled E-Voting," IEEE Software, vol. 35, pp. 95–99, July 2018.
- [3]. M. Pawlak, J. Guziur, and A. Ponsizewska-Maran´da, "The Blockchain Voting Process: Auditable Blockchain Voting System," in Lecture Notes in Data Engineering and Communication Technologies, pp. 233–244, Springer, Cham, 2019.
- [4]. B. Singhal, G. Dhameja, and P. S. Panda, "How Blockchain Works," in Beginning Blockchain, pp. 31–148, Berkeley, CA: Apress, 2018.
- [5]. Agora, "Agora Whitepaper", 2018.
- [6]. R. Perper, "Sierra Leone is the first country to use blockchain during elections - Business Insider", 2018.
- [7]. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," Tech. Rep., 2008.
- [8]. G. Wood et al., "Ethereum: A Secure Decentralized Generalized Transaction
- [9]. R. Perper, "Sierra Leone is the first country to use blockchain during an election - Business Insider", 2018.
- [10]. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," tech. rep., 2008.
- [11]. G. Wood et al., "Ethereum: A secured decentralised generalised transaction ledger," Ethereum project yellowpaper, vol. 151, pp. 1–32, 2014.
- [12]. S. Landers, "Netvote: A Decentralized Voting Platform - Netvote Project - Medium", 2018.
- [13]. P. McCorry, S. F. Shahandashti, and F. Hao, "A Smart Contract for
- [14]. Boardroom Voting with Maximum Voter Privacy," in Lecture Notes in Computer Science, ch. FCDS, pp. 357–375, Springer, Cham, 2017.
- [15]. Z. Brakerski and V. Vaikuntanathan, "Efficient Fully Homomorphic Encryption from (Standard) LWE," SIAM Journal on Computing, vol. 43, pp. 831–871, Jan 2014.