सीएसआईआर- एम्प्री
CSIR-AMPRI

# Blockchain-Based Chain Of Custody: A Secure Digital Evidence Framework For Digital Forensics Investigation

**Febin Prakash* & Harsh Sadawarti**

*CT university, Punjab, India*

**Abstract** : The rapid expansion of digital space has led to a rise in cybercrime, stressing the importance of actual evidence in building a relationship between suspected individuals as well as cybercrime. It is necessary to have a chain of custody (COC) for any evidence object, which is a document of movement and possession. As technology progresses, the safety of digital evidence (DE) becomes increasingly important in forensic investigations. When it comes to COC, DE presents its own set of challenges. Existing database systems aren't capable of understanding the requirements for the accuracy as well as the validity of DE. Blockchain-based COC is a system for preserving as well as analyzing evidence in digital forensics (DF). CA blockchain is a data format that enables all users in a distributed network of computers to build a digital ledger for storing and maintaining transactions. Blockchain (BC) creates an undeniable audit trail by encrypting the process of storing and managing network transactions. This study develops a basis for implementing DE authentication, integrity, and privacy, making it a reliable solution that retains evidence authenticity and ensures its permissibility among various stakeholders involved, like law enforcement agencies, solicitors, and forensic experts.

**Keywords**: Blockchain; Forensics; Digital Evidence; Chain of Custody; Cyber Forensic; Digital Forensics Investigation.

## Introduction

As the devices connected to the internet grow, DF has risen to include all of the different technologies we use every day (Navarro-Ortiz et al., 2018). DF is a well-established skill domain in cybersecurity, and it is a vital component of an incident response strategy using electronic information. The primary purpose of DF is to perform digital analyses within a legislative structure in response to illegal acts involving computer technology. In a civil or criminal process, the objective is to prove or reject a hypothesis. In this case, eDiscovery could be applied to resolve disagreements among various commercial parties. Skilled and experienced investigators collect, assess, and recreate incidences as well as actions using forensically appropriate methods (extensively analyzed and validated) to help describe what happened in support of a case (Daryabar and colleagues, 2017). The scope of DF is continually expanding. Specialists in mobile phone and computer forensics, onsite (scene of the crime) examinations, records in call data, demand orders, forensic readiness plans, storage and retrieval, and audio-visual forensics are necessary to establish a successful team. Because of the highly integrated cyber-physical environment we operate in, non-electronic information is also included.

DE is defined as any electronic information comprising correct information to support an event hypothesis. DE's scope is continuing to expand,

implementing both existing as well as emerging technologies such as computers, smartphones, networks, and memory (Ali et al. 2022). In DE, the ease with which it can be reproduced or disseminated, as well as the ease with which it can be changed or damaged are all factors. DE is also time-sensitive. There is also the convenience of transferring DE between countries. Assessing DE is thus more difficult than assessing physical evidence (Prayudi & Sn, 2015). Digital evidence includes images, texts, videos, and device records. This research proposed a basis for implementing DE authentication, integrity, privacy, and a secure solution that ensures evidence integrity and permissibility among various stakeholders, such as law enforcement agencies, solicitors, and forensic

*Corresponding author:(E-mail: febin18002@ctuniversity.in)

experts (Kumar et al., 2021).

Due to the widespread availability of image manipulation software and the growing prevalence of digital photography, digital picture forgeries are becoming increasingly prevalent. It's impossible to determine whether the photograph is genuine or has been modified. A portion of a photograph can be removed, a portion of the photograph can be obscured, or the photograph can be altered so that the image data is displayed improperly. These issues affect the reliability of digital photographs (Patel et al., 2017). A variety of methods for detecting image deception are carefully discussed. They are divided into active algorithms (AA) and passive algorithms (PA). The AA entails putting a watermark on the picture. Methods for passive forgery identification look at evidence left on the picture after many picture processing stages. Additionally, it can be used to determine the amount and position of fraud in a photograph (Varkey & Nair, 2018).

Tian et al. developed a secure DE framework based on BC technology in 2019. It includes a loose-coupling format which preserves both the evidence as well as the evidence data in different locations. The researchers (Widatama, Prayudi, and Sugiantoro, 2018) used the RC4 cryptographic technique to encrypt the XML layout on the digital COC data storage. No database management system (DBMS) must use this XML format, which is simple enough for non-experts to understand. DE cannot be accepted in court since the information is accessible to everyone.

Furthermore, unlike earlier BC-based picture forensics systems, which used conventional hashing to validate the BC validity, the proposed method uses fuzzy hashing to properly manage evidence of object alterations produced by malicious as well as suspicious attacks. Whenever the correlation between the two blocks exceeds 95%, the block is viewed separately (Lone et al., 2019). This study examines the methods used in the study as well as the findings that were drawn from it.

## Methodology

This section shows how to handle defects in the DE for multiple copies of a similar document (unpredictability about the integrity).

All picture forensic-capture technologies are included in the data-gathering step. Information from hard drives, RAM data, operating systems, application logs, network packet captures, as well as smartphones must be collected in accordance with forensic standards by the expert during this phase of the investigation. Regarding the certificate of authenticity, BC technology can record tamper-proof evidence, particularly when paired with fuzzy hashing. Because traditional hash methods are ineffective in this scenario, forensics experts can properly solve the issue of authorized modification of DE by utilizing fuzzy hash functions.

The effectiveness of the proposed system has been validated for use in picture forensics. This technique can convert a set of data types such as audio, video, photos, and documents.

The fundamental procedure of the proposed framework is depicted below.

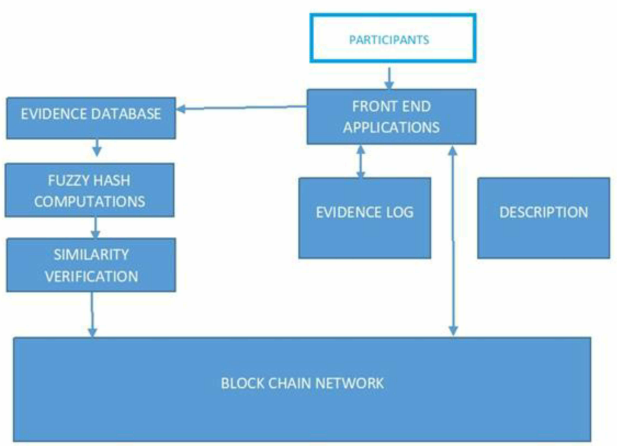


*Fig. 1: Proposed Model*

**The following factors influence the selection of photos in the application:**

(1) Several occurrences in DF specialists' work are associated with picture forgery, as images of signs and cheques constitute the majority of transactions made.

(2) The development and availability of advanced image analysis application programs and computer technologies have made manipulating digital photos incredibly simple. A comprehensive study is required to assure digital photographs' validity, integrity, accuracy, and origin.

(3) Images have been used in highly specialized domains like forensic science, astronomy, medicines, and surveillance. The researcher does not affect the evidence, although minor changes made inside some programs, such as contraction, may be altered. Even though a single aspect of the input is altered, cryptographic hash methods' pseudo-random nature prevents identical files' subsequent detection. When working in DF, it is necessary to use a hash function that doesn't keep

file commonalities (for example, various versions of the same file).

## Results

Performance is perhaps the desirable attribute of any problem-solving activity. So, solutions based Blockchain are no exception. During the course of this analysis, the Hyper Ledger Caliper was utilized in order to determine the overall effectiveness of the proposed system. Performance measures such as transaction per second (TPS) and transaction delay can be used to compare different block chain networks in terms of their ability to meet a set of use cases(moment spent between the time a transaction was made and the time when it was recorded in the BC). The code was written using Python 3.6 software. Caliper's two-organization-one-peer as well as three-organization-one-peer network models were applied in the 1st round of assessment to test our prototype with 4 customers using Caliper's two-organization-one-peer as well as three-organization-one-peer network models.

Since they had a direct effect on the state of the BC, this study made a test document that looked at two important parts of our approach, evidence creation and transmission. 10 rounds of evaluation were conducted with a variety of transaction quantities and transmit transaction rates to determine the best configuration. In order to obtain average values for vital aspects with the lowest probability of error, multiple tests have to be conducted. As per the results of the performance investigation, the prototype's throughput reaches a max before diminishing as the transmit rate (TR) increases. Both two-organization-one-peer and three-organization-1-peer network topologies have attained the highest throughputs, with fifteen TPS and ten TPS. The outcomes, however, reveal that increasing the number of colleagues has an influence on the throughput of the prototype. It is typical of hyper ledger-based coalitions BC. The evaluation of how well the proposed system would work is shown in the table below.

***Table 1: Performance Valuation of proposed system***

| R | SR (tps) | MxL(s) | MnL (s) | AL (s) | TP (tps) |
|---|---|---|---|---|---|
| 1 | 6 | 0.84 | 0.67 | 0.88 | 5 |
| 2 | 11 | 1.17 | 0.70 | 0.67 | 9 |
| 3 | 16 | 1.15 | 0.71 | 1.56 | 13 |
| 4 | 20 | 2.87 | 0.89 | 1.57 | 14 |
| 5 | 25 | 4.79 | 0.77 | 2.78 | 16 |
| 6 | 31 | 5.11 | 1.09 | 4.78 | 15 |
| 7 | 36 | 11.90 | 1.89 | 5.64 | 14 |
| 8 | 41 | 22.68 | 8.25 | 16.89 | 8 |
| 9 | 45 | 12.78 | 2.67 | 8.34 | 13 |
| 10 | 48 | 13.89 | 8.57 | 11.34 | 15 |

*R – Round, SR – Send Rate, MxL – Max Latency, MnL – Min Latency, AL – Avg Latency, TP – Throughput

It is shown in Table 1 the latency as well as throughput for a variety of 2- and 3-organization 1-peer network configurations. The prototype's throughput reaches a max. value during the performance assessment, and afterwards gradually declines as the transmit rate rises.

Block generation was researched in the 2nd test phase and the number of blocks formed by every node was calculated. This value indicates whether or not each BC node has a fair probability of creating blocks. The cumulative (Cum.) proportion of blocks generated by x nodes can be seen in the following graph (fig. 2). Here, 'k' represents ₹the number of node names₹. The line is more likely to be straight if the weight is evenly distributed. The curve begins a significant ascent when k is equal to one.

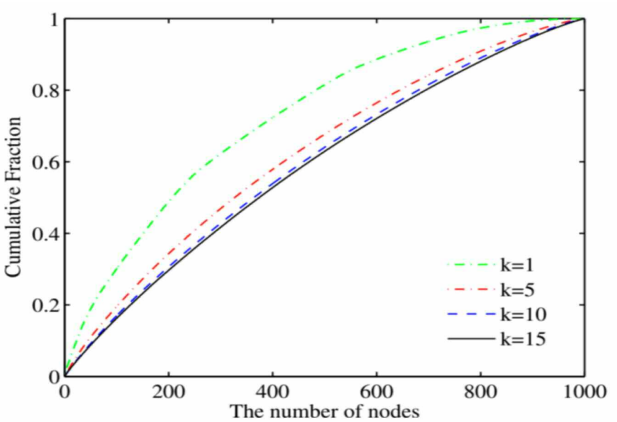


*Fig. 2: Cum. Dispersal of Blocks*

The final set of experiments used a topology with 1,000 nodes to determine the BC dimensions when different block dimensions were used. The graph below displays the size of the BC as a function of the number of blocks. The mixed BC is utilized to compute the max, mean, as well as minimal BC dimensions, while the entire size of the BC is computed using a common scenario whereby all nodes possess the whole BC.

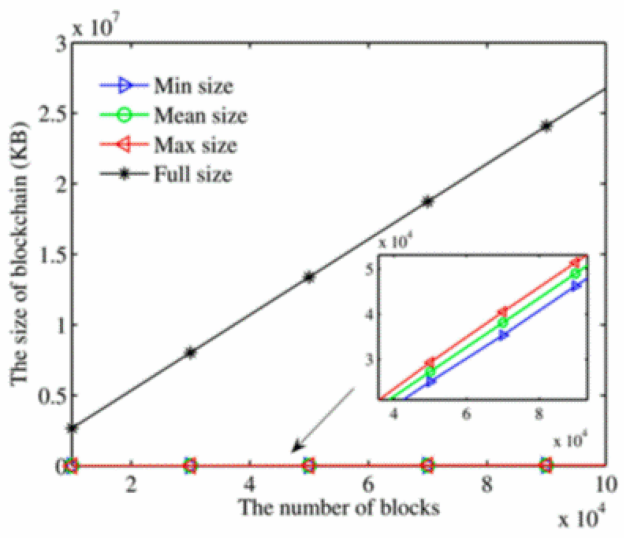


Fig. 3: Dimension of the BC

In the standard BC, the mixed BC is a subset. When the number of blocks increases exponentially in all four cases, this supports the theoretical theory. It was discovered in this research that MRSH–method v2's of searching for ₹illegal₹ documents took a lot longer than this approach's method of searching for the 100 ₹illegal₹ documents contained verbatim in the hard drive image, as well as the 40 ₹illegal₹ documents found in the picture. More than 4,000 ₹known–illegal₹ photos were included in a collection that also included 140 additional images. Of the 4000 ₹illegal₹ photographs analyzed by MRSH–v2, 100 were found to be similar to those in the database, while the other 40 were not.

The main indicator was how long it took to finish the whole method that comprised building the tree, searching for it, and analyzing the leaves in pairs. The running time is shown in the diagram below.

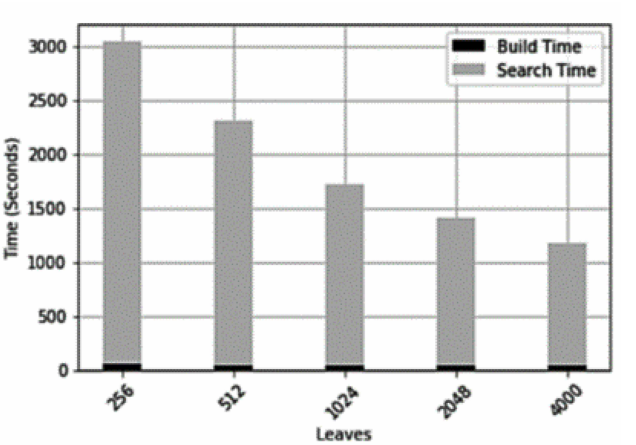


*Fig. 4: Running Time*

The tree began with a sample of 4000 "illegal" images and was then enlarged to include all of the images in the larger collection. When calculating "Search Time," we include the time it takes to look through the entire tree and compare the leaves. In order to get the fastest execution time, more leaf nodes were used. Between the start and finish of the race, there were 1197 seconds that is "all–against–all" comparisons take 49 percent less time than pair–wise comparisons. Because the paired technique doesn't scale, this difference is likely to become even more obvious with larger datasets.

Fuzzy Hashing (FH) was used in this study to account for changes in the evidence items.₹ Piecewise and Rolling Hashing₹ are both parts of FH (RH). CTPH is called a ₹grey hash type₹ because it can tell if two files are almost the same, but other hashing methods wouldn't be able to tell. With RH, input context determines how long traditional hash strings will be generated into. It is possible to build a checksum for a complete image using Piecewise Hashes (PH) and to get over this limitation, they divide the image into preset sections and hash each one. The final hash sequence is the created values. In this study, FH uses PH to keep data similar. PH also ensures data integrity by guaranteeing that one hash segment is empty during forensic imaging. Memory storage of the proposed system uses idle hard drive space from users to store data. Decentralized infrastructure can overcome several difficulties with centralized cloud storage.

In terms of forensics, the proposed process is complex because it can be attacked in both ways. You can hide information with anti–blacklisting and anti–whitelisting. Attackers modify files such that fuzzy hashing doesn't recognize them as being the same. As far as humans are concerned, there is no discernible change between the original and the edited version.

When a file is modified successfully, it is marked as ₹unknown, not bad.₹ This way to stop blacklisting changes one bit within every chunk as well as keeps track of trigger points. Alter the trigger so that the Hamming distance tells how big each change is. Each building block has a Hamming distance, and triggering can be changed with a ₹one-bit modification.₹ Active opponents must change one bit each time they meet. There are more places where the Hamming distance is short, so 100 more changes are needed. A whitelisted file's hash value must be used to change a bad file such that its hash value matches that of a whitelisted file in order for anti-whitelisting to work. An attack's original and altered forms are indistinguishable to humans. Using this method, a given signature can be created by creating legal trigger sequences then inserting zero-strings. If a file's hash value can be altered in any way, it will no longer be useful. All active trigger sequences are erased when an adversary is active. In the second step, he has to replicate the white-listed file's triggering behavior, which requires a number of system modifications.

**Conclusions**

These processes depend on the reliability and reliability of DE to manage the COC in a unified manner (or chain of evidence). Fuzzy cryptographic hash algorithms in BC technology are compared to regular cryptographic hash algorithm methods to examine how good they are at protecting the integrity of DE in picture analysis for determining commonalities. We developed and evaluated a forensic chain model prototype using a hyper ledger component. Because of its capacity to deal with COC-related unpredictability and keep a realistic workload, the fuzzy hash-based BC was shown to be an excellent support for the COC method in the performance evaluation results. The suggested framework performance will be tested in the long term when working with multiple digital forms of evidence.

## References


[1] Ali, M., Ismail, A., Elgohary, H., Darwish, S., & Mesbah, S. (2022). A Procedure for Tracing Chain of Custody in Digital Image Forensics: A Paradigm Based on Grey Hash and Blockchain. Symmetry, 14(2), 334.

[2] Prayudi, Y., & Sn, A. (2015). Digital chain of custody: State of the art. International Journal of Computer Applications, 114(5).

[3] Navarro-Ortiz, J., Sendra, S., Ameigeiras, P., & Lopez-Soler, J. M. (2018). Integration of LoRaWAN and 4G/5G for the Industrial Internet of Things. IEEE Communications Magazine, 56(2), 60-67.

[4] Daryabar, F., Dehghantanha, A., & Choo, K. K. R. (2017). Cloud storage forensics: MEGA as a case study. Australian Journal of Forensic Sciences, 49(3), 344-357.

[5] Patel, J. J., & Bhatt, N. (2017). Review of digital image forgery detection. Int. J. Recent Innov. Trends Comput. Commun, 5(7), 152-155.

[6] Varkey, A., & Nair, L. (2018). Robust image forgery detection and classification in copy-move using SVM. Int. J. Adv. Res. Trends Eng. Technol, 5(12), 89-93.

[7] Tian, Z., Li, M., Qiu, M., Sun, Y., & Su, S. (2019). Block-DEF: A secure digital evidence framework using block chain. Information Sciences, 491, 151-165.

[8] Widatama, K., Prayudi, Y., & Sugiantoro, B. (2018). Application of RC4 Cryptography Method to Support XML Security on Digital Chain of Custody Data Storage. International Journal of Cyber-Security and Digital Forensics, 7(3), 230-238.

[9] Kumar, G., Saha, R., Lal, C., & Conti, M. (2021). Internet-of-Forensic (IoF): A block chain based digital forensics framework for IoT applications. Future Generation Computer Systems, 120, 13-25.

[10] Lone, A. H., & Mir, R. N. (2019). Forensic-chain: Blockchain based digital forensics chain of custody with PoC in Hyperledger Composer. Digital investigation, 28, 44-55.