International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN :2147-6799 www.ijisae.org Original Research Paper

Designing A Secured Framework for the Steganography Process Using Blockchain and Machine Learning Technology

Ayushi Chaudhary[1*], Ashish Sharma[2], Neeraj Gupta[3]

Submitted: 15/10/2022 **Revised:** 21/12/2022 **Accepted:** 20/01/2023

Abstract: Recently, Blockchain technology has become popular. Through a highly secure, decentralized system enabled by this technology, anyone can transact securely without the need for a middleman. Machine learning can assist in addressing many of the constraints that Blockchain-based systems have in addition to their own potential. Machine learning and Blockchain technology can be combined to produce effective and practical outcomes. The infrastructures, resources, end devices, and applications in communications and networking systems are becoming increasingly more complicated and heterogeneous recently due to the rapid growth of information and communication technology. Additionally, the enormous amount of data and numerous endpoints could pose significant security, privacy, delivery of services, and network management difficulties. The combination consideration of Blockchain and Machine Learning is necessary to accomplish decentralized, secure, intelligent, and effective network operation and management (ML). This article will explain Blockchain technology and examine how machine learning capabilities can be combined with a Blockchain-based system to build a stego cryptography to secure the data during communication over the network.

Keywords*: Blockchain, stego cryptography, machine learning, decentralization.*

1. Introduction

Blockchain is expected to be the "future of financial and cybersecurity," with the potential to "revolutionize apps and reshape the digital economy," according to experts [1]. Blockchain has a lot of potential for re-establishing "trust" in society by allowing for coordination without relying on third parties [2]. If this is the case, developing Blockchain into a functional, digital institutional infrastructure with transparent governance, security, and operating standards is critical. Data is stored in a distributed ledger in Blockchain. Participants in the Blockchain network can create, read, and verify transactions recorded in a distributed ledger using Blockchain technology, which provides integrity and availability. It does not, however, allow for the deletion or modification of transactions or other information kept on its ledger. Cryptographic primitives and protocols, such as digital signatures and hash functions, support and safeguard the Blockchain system. These primitives ensure that transactions recorded in the ledger are protected from tampering, authenticated, and non-repudiated. Decentralization, autonomy, integrity, immutability, verification, fault-tolerance, anonymity, auditability, and transparency are all desired qualities of Blockchain in a trustless environment, which has received a lot of academic and industrial attention in recent years [3][5]. This paper presents and explains a revolutionary steganography-based Blockchain technique and Machine Learning(ML). For updating and exchanging the information, stego pictures with hash data and Blockchain technology with ML are used. Confidential data shared on the network improves confidentiality and ensure the integrity of data. The rest of the paper is organized as follows: Section 2 gives the overview and literature review on the works done on steganography using Blockchain and Machine Learning, section3 explains the Stegno-Cryptography using Blockchain and Machine Learning Section4 explains the Blockchain technology with ML in detail including stegano cryptography for the Blockchain. Section 5 explores how Blockchain and Machine Learning can benefit in the security privacy with the proposed model, and finally, section 6 concludes the work with a summary.

ayushichaudhary11@gmail.com[1],ashish.sharma@gla.ac.in[2], neeraj.gupta@gla.ac.in[3],*
GLA UNIVERSITY Mathura.

2. Overview of Blockchain and Machine Learning:

Blockchain-based systems combine cryptography, public key infrastructure, and economic modeling to accomplish distributed database synchronization through peer-to-peer networking and decentralized consensus. For the better part of a decade, Blockchain technology (BCT) [6] has gotten a lot of interest throughout the world. Researchers' attention has switched to the technology's findings after it was first coined. The term "digital currency" or "cryptocurrency" was coined by Satoshi Nakamoto. These

cryptocurrencies are built on decentralized networks, and Blockchain is the technology that powers them.

Blockchain is a decentralized [4], immutable database that makes it easier to track assets and record transactions in a corporate network. A tangible asset (a house, car, cash, or land) is different from an intangible asset (intellectual property, patents, copyrights, branding). On a Blockchain network, virtually anything of value may be recorded and traded, lowering risk and costs for all parties involved.

The following are the essential components of a Blockchain [h1]:

a) Technology based on distributed ledgers:

The distributed ledger and its immutable record of transactions are accessible to all network participants. Transactions are only recorded once with this shared ledger, eliminating the duplication of effort that is common in traditional commercial networks.

b) Immutable records:

After a transaction is recorded to the shared ledger, no participant can edit or tamper with it. If a mistake is found in a transaction record, a new transaction must be entered to correct the problem, and both transactions are then visible.

c) Contracts with intelligence:

A collection of rules called a smart contract is stored on the Blockchain and executed automatically to speed up transactions. A smart contract can specify requirements for corporate bond transfers, as well as payment terms for trip insurance.

Machine Learning:

A. L. Samuel introduced the term "machine learning" (ML) in 1959, and it is defined as "the field of study that offers computers the ability to learn without being explicitly programmed" [20]. E. Tom Mitchell later provided a better definition of machine learning (ML) as "a computer program said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." Different Machine Learning algorithms are used to process the data. Training and testing processes are frequently included in an ML framework's normal workflow. Before moving on to the next step, the raw data are first pre-processed during the training phase. Following that, this data's features and patterns can be retrieved and processed for the specified tasks.

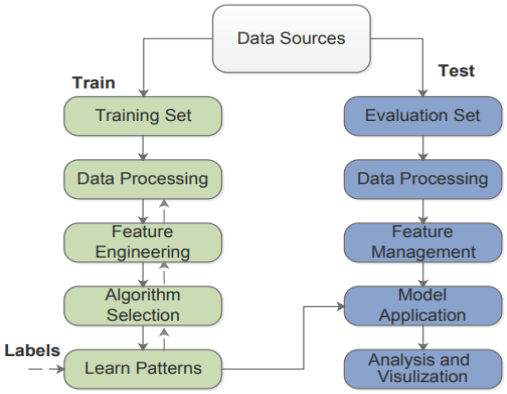

Fig. 2: Machine Learning Model Workflow

Need for the Blockchain:

Duplicate record keeping and third-party validations waste a lot of time in operations. Fraud and cyberattacks can make record-keeping systems susceptible. Data verification might be slowed by a lack of openness. And, with the advent of the Internet of Things, transaction volumes have skyrocketed. All of this slows commerce and depletes the bottom line, indicating that we need to find a better solution.

2. Literature Review

In his Ph.D. thesis from 1982, Chaum was the first known person to propose a Blockchain-like protocol [7]. Haber and Stornetta described a cryptographically secure chain of blocks in 1991 [8]. Bayer et al. [9] added Merkle trees into the design in 1993. Szabo [h1] created "bit gold," a decentralized digital currency method, in 1998. Nakamoto introduced Bitcoin, a peer-to-peer electronic cash system, in 2008 [10]. In 2008, the word "Blockchain" was coined to describe the distributed ledger that underpins Bitcoin transactions [11].

Zhao et.al. [13] propose a data-hiding strategy based on Blockchain technology for digital video data protection. The amalgamation of Blockchain and data hiding techniques is introduced to improve the reliability of video encrypted data. The has value is added to each block with

the traceable hash value record. The author proposes a DCT-based data concealing scheme that interacts with the Blockchain record, where the external chain is used as the controller for the privacy data into carrier video, to improve the level of security of private data such as copyright linked to digital video. The author concludes protection for data-hiding servers, particularly for decentralized server distribution, must be considered in the future.

Nelaturu et.al.[14] propose public and private Blockchain are explained and also propose depth knowledge of fundamental principles required for comprehending technology. The author fills the gap between Blockchain and Fintech applications. Different Blockchain classification technologies are explained to address the issues of privacy and security in the Fintech online services.

Mohsin, Ali H.et.al. [15] the author presents and explains a new steganography-based Blockchain technique in the spatial domain. The particle swarm optimization (PSO) algorithm with hash function is applied to hide the COVID-19 medical dataset which provides high-efficiency capacity and good image quality. The author applies three stages in the proposed method as the pre-hiding stage which embeds the host image, the COVID-19 medical dataset hiding using PSO and hash function, and finally stego images with Blockchain technology are applied to improve the confidentiality level and protecting and securing the medical dataset during the transmission on the network.

Elisa et.al.[16] in their study presents a framework for a decentralized e-government peer-to-peer (p2p) system based on Blockchain technology that may protect data security and privacy while also enhancing public sector trust. Blockchain technology is used to ensure high security and privacy by preserving decentralized systems where transactions are not under third-party control. To avoid cyber-attack such as malware, denial of service (DoS), and distributed denial of service attacks (DDoS) due to the failure in the centralized management and system validation, Blockchain technology is proposed. The author proposes the prototype with theoretical support and a quantitative analysis of privacy and security.

Liu, Si et.al. [17] in their study explores the features of Blockchain that benefit steganography. To safeguard the privacy of Blockchain transaction data, this study first separates each transaction into two parts: sensitive data and basic data, and then encrypts and hides the sensitive data in HEVC video. The results of the experiments reveal that this HEVC video steganography algorithm effectively increases the privacy data embedding capacity while maintaining high visual quality.

Sarkar el. al.[18] the integrity of the stego image is a major worry, as an intruder could intercept the image and disrupt the secret connection. The author overcomes this issue by introducing a Stego-chain-based Blockchain platform that employs a steganographic-based efficient approach to enhance the Robert Edge Detection method. The author proposes a privacy protection Blockchain transaction method for HEVC video steganography. After encryption, the private data is embedded into the 4 4 luminance QDST blocks, and the suggested steganography approach has superior embedding capacity and visual quality thanks to the usage of the matrix encoding technique. The proposed method's viability and superiority are demonstrated by experimental findings.

Alizadeh et.al.[19] Using biometrics, machine learning, a decentralized hash table, and Blockchain technology, the study illustrates a decentralized smart identity strategy for video conferencing applications. In this study, the immutability and traceability features of distributed ledger technology with the unlimited storage capability of distributed hash tables to increase the system's storage capacity and immutability are applied. Using an open Blockchain and IPFS, this study demonstrated a machine learning-based facial recognition system. The authors discuss support for decentralized web hosting and discussion data sharing in a DHT or IPFS.

Liu, Yiming et.al. [20] give a survey of the literature on Blockchain and machine learning technology. Discover the overview, advantages, and applications of merging Blockchain with ML, among other key factors. The combination of Blockchain with ML, which is becoming a crucial solution to enable intelligent, safe, and decentralised sharing of data and models as well as the effective operation of communications and networking systems, is covered in this paper.

3 Stegno-Cryptography using Blockchain and Machine Learning:

Traditional image data of steganography which can be in the form of pictures, audio, or video encrypts secret data into the cover image before sending the stego-image to the intended recipient through an insecure channel. The attackers try to detect the channel but are unable to discern any signs of the hidden data because both the cover and stego images appear to be indistinguishable. The growth of the Internet and communications has made the exchange of sensitive data among many parties

quite simple. However, insecure channels present problems with secrecy, authenticity, and integrity, among other things. Cryptography and Steganography are two generally accepted solutions used by researchers to address these issues for a long time. Although encrypted material attracts attention and reveals its importance when intercepted, cryptography, the process of changing sensitive data into a non-readable form, offers safe data transmission. Steganography, on the other hand, the process of hiding confidential data, reveals no information about this covert communication. Both spatial and transformed domains are used to implement steganographic methods. The integrity of the stego image, on the other hand, is a major worry, as an intruder could intercept the image and disrupt the secret conversation. To overcome this problem, "Stego-chain" with a Blockchain-based platform that employs an efficient steganographic approach.

There have been considerable advancements in developing Blockchain-based authentication systems that enable excellent security in a decentralized network, according to Blockchain research. As an open decentralized technique of creating confidence, Blockchain technology is introduced as the core mechanism for crypto-currency in Bitcoin [12]. Blockchain is a decentralized platform for exchanging information among nodes that is a public record of information acquired through a network, represented as a chain of blocks. Using cryptographic hashing, these blocks are linked to one another. By protecting the integrity of the data, it makes it tamper-proof. Individuals or groups of people that do not trust each other to coordinate and participate in a cohesive decision-making process but want to share information on a shared platform can benefit from Blockchain technology. Blocks, nodes, and miners are the three main components that make up a Blockchain.

Blocks: The chain is made up of numerous blocks, each of which includes two parts: a header and data. A block's header typically contains the hash of the previous block, the hash of the current block, and a nonce, which is a 32-bit whole number. The data section includes the information about the transactions that have been completed. In the proposed framework the image, video, or audio is considered as a type of data in which the data is hidden. The data is first encrypted using a shared secret key, then fragmented into chunks of frames (i.e., m n pixel matrix), and last the frames are communicated with their signatures. Each frame is represented as a transaction when combined with its signature on the internet. Each block of a Blockchain is seen as a collection of confirmed transactions in this sense. Every block is linked in chronological order and connected to each other. The framework with the data in the block is shown in figure 1.

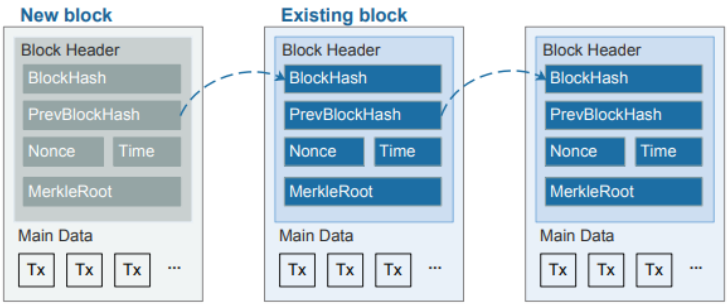

Fig. 1: Framework of Blockchain

The sender's address: This contains a 256-bit whole number that is used to recognize the sender on the decentralized network.

Receiver's address: The decentralized network uses a 256-bit whole number to uniquely identify the receiver.

Block index: It specifies the block's position within the Blockchain. The starting block is also known as the Genesis block has the index "00", with the "10" as the second index has the index '00,' and so on.

Timestamp: This is a 32-bit whole number that is collected when a block is created. A random timestamp, for example, looks like this: 2020–05–26 12:35:467.

When a block is constructed, a 32-bit whole number is generated at random, which is then used to generate a block header hash.

Nonce: To find a hash value lower than the value of a target, once starts at zero and increases to infinity. Unless it is mined, the data in the -block is signed and irrevocably linked to the nonce and hash.

Current hash: The nonce is paired with a 256-bit integer. It has to start with a lot of zeros.

Previous hash: The hash of the previous block is represented by a 256-bit integer. The previous hash is treated as a ten-second string in the Genesis block.

Decentralized Hash Table:

Based on distributed hash table technology [22], IPFS decentralized system without a single point of failure. In order to reduce excessive file duplication, IPFS, a decentralized storage technology, assigns a distinct hash to each file that is stored. The unique hash address allows the user to retrieve the appropriate file. Without memory capacity restrictions, IPFS aids in the storage of decentralized data [23]. Blockchain's inherent decentralization offers the core protocols required for decentralized machine learning (ML) algorithms that take security and privacy into account. By using Blockchain, machine learning (ML) might train and make decisions on local devices in decentralized and distributed networks. Every stage in the lifespan of ML models can be improved with Blockchain technology, from training to optimization.

3. Methodology

From the related works, steganography with Blockchain is applied either taking the image or audio or video. The research gaps are protection for data-hiding servers, particularly for decentralized server distribution are not considered, and the security level of protecting the data is done by taking digital video but not on images and audios. The malicious members may attack the Machine Learning based system over the network systems during the communication. To deal with this issue Blockchain technology with decentralized and distributed ledger is used. Blockchain and machine learning (ML) are viewed as potential technologies to facilitate decentralized, secure data and model sharing as well as intelligent network operation and administration. The Blockchain features such as decentralized, cryptographic, and transparent provide security and privacy to ML during the decision-making process with an assurance of records have not been tampered.

4. Proposed Model

The sender select the images extracted from the input either from video or audio or text images using the Machine Learning Algorithm such as support vector machine, recurrent neural network, convolution neural network (CNN), and decision tree (DT). Machine Learning consists of 1: dataset, stage 2: pre-processing, stage 3: normalisation, stage 4: training, and stage 5: testing. The dataset stage involves choosing a benchmark dataset that includes various attack scenarios. The data set is divided into training and test sets, and each row is classified as an attack or benign during the pre-processing stage. The various aspects of the data collection are normalized during the normalisation stage.

The extracted image is called the cover image(CI). The sender side computes the edge image (EI) over the cover image selected by the sender and embeds the images into pixels of CI. To ensure the security of communication end-to-end encryption is done with the shared secret key(α). The encrypted image is called the stego image. The sender extract m x n pixel matrices from the encrypted stego image and constructs the list of frames to be transmitted. The signature key which is a private key of a sender is computed for each frame. The decentralized network then broadcasts each transaction to all active nodes for verification. Each node selects the transaction for verification one by one by using the sender’s signature. The miner then verifies each transaction individually, groups several verified transactions into a block, creates a distinct hash by resolving a challenging puzzle, and attempts to chain the blocks together. A reward is given to the miner who validates the block and solves the puzzles first. The block index, timestamp, difficulty level, nonce, previous hash, and current hash are all included in the block header along with the sender's and receiver's addresses. The block containing the sender's transaction is added to the Blockchain once the peer-to-peer network has reached consensus and has verified all transactions. The virtual point as a reward is been used in order to evidence proof of work(PoW), which represents the reliability of the miner.

The end receiver receives all validated blocks, retrieves the complete batch of frames from the Blockchain, and then reconstructs an encrypted stego image (ESI). As the ESI blocks are verified by the intermediate nodes, these frames are free from tampering. As a result, using the shared secret key, the receiver will be able to retrieve SI from

ESI. Additionally, the receiver calculates EI from the 3-MSBs of each pixel in SI and extracts the next 'x' or 'y' bits to recreate SD.

Two important aspects, namely the payload and the stego-image quality, have been taken into consideration in order to summarise the outcomes of the suggested approach. To calculate the findings, 50 benchmark images of dimension 128 x128 are used (Weber, 1997; Fei-Fei et al., 2004) as shown in Figure 4. The average amount of hidden bits that a stego-pixel can produce is the payload. The unit of payload in this instance is bit per pixel (bpp). Instead, the peak signal to noise ratio (PSNR), mean squared error (MSE), structural similarity index (SSIM), and universal image quality index (UQI) have each been used to evaluate the quality of stego-images. Equation (1) defines the PSNR (in dB).

$$PSNR = \log_{10}(MAX1)/\sqrt{MSE} \quad \text{----------} \quad (1)$$

This MSE value is 255 since an image (I) can only be represented by 8 bits per pixel. The average squared difference between each original pixel and its warped counterpart is how the metric MSE is represented. It is calculated by adding up the squared differences

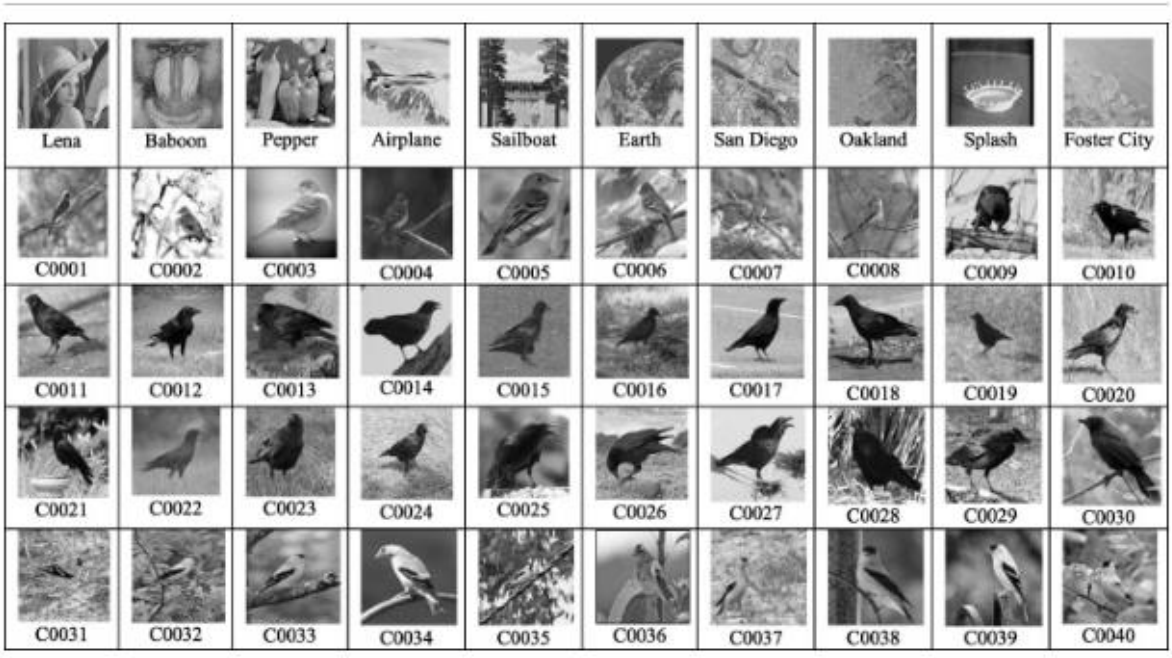

Fig. 4: Benchmark images of dimension 128 x128 used for stegnocrytography

between each pixel, then dividing the result by the total number of pixels. The MSE can be defined using equation (2) in terms of the concept of steganography.

$$MSE = \frac{1}{W\,X\,H} \sum_{i=1}^{w} \sum_{j=1}^{H} (Xij - Yij)2 \quad \text{----------} \quad (2)$$

The stego image is evaluated through existing methods using PSNR value, and mean square error (MSE).

4 Blockchain and Machine Learning can benefit the security privacy:

From the previous explanation and proposed model, Blockchain can be advantageous with Machine Learning with regards to data and model sharing, security and privacy, decentralized intelligence, and reliable decision-making. The data communicated by extracted using relevant features using the Machine Learning technique need security and privacy. The present ML techniques usually adopt a centralized architecture which may lead to error-prone hacking. Moreover, the trained data usually involves a huge amount of important information, if proper care is not taken then the data may lead to privacy concerns. To protect the ML models from hackers the Blockchain plays an important role.

5. Conclusion

The centralized feature of Blockchain technology provides makes the ML algorithms more secure, and privacy. By applying Blockchain technology, ML can learn, train, and can make decisions making on local devices in decentralized and distributed networks. In this study Blockchain-based stegno cryptography by applying Machine

Learning technique was proposed. This proposed system improves the efficiency in terms of data protection while sending the information through a network. The proposed method of hiding the input data includes data protection on chain in the block, the relevant data can be extracted using the using best machine learning algorithms. As shown in the proposed work, the combination of Blockchain with Machine Learning technique in hiding the data behind the selected input which can be text image, video or audio can improve the protection level of secrete data. The feature of Blockchain gives high accuracy level during the communication of information

References

[1] Underwood, Sarah. "Blockchain beyond bitcoin." *Communications of the ACM* 59, no. 11 (2016): 15-17.

[2] Shahaab, Ali, Ross Maude, Chaminda Hewage, and Imtiaz Khan. "Blockchain-a panacea for trust challenges in public services? A socio-technical perspective." *The Journal of the British Blockchain Association* (2020): 14128.

[3] Ali, Muhammad Salek, Massimo Vecchio, Miguel Pincheira, Koustabh Dolui, Fabio Antonelli, and Mubashir Husain Rehmani. "Applications of blockchains in the Internet of Things: A comprehensive survey." *IEEE Communications Surveys & Tutorials* 21, no. 2 (2018): 1676- 1717.

[4] Troncoso, Carmela, Marios Isaakidis, George Danezis, and Harry Halpin. "Systematizing decentralization and privacy: Lessons from 15 years of research and deployments." *Proceedings on Privacy Enhancing Technologies* 2017, no. 4 (2017): 404-426.

[5] Lin, Iuon-Chang, and Tzu-Chun Liao. "A survey of blockchain security issues and challenges." *Int. J. Netw. Secur.* 19, no.5 (2017): 653-659.

[6] Komalavalli, C., Deepika Saxena, and Chetna Laroiya. "Overview of blockchain technology concepts." In *Handbook of Research on Blockchain Technology*, pp. 349-371. Academic Press, 2020.

[7] Chaum, David Lee. *Computer Systems established, maintained and trusted by mutually suspicious groups*. Electronics Research Laboratory, University of California, 1979.

[8] Haber, Stuart, and W. Scott Stornetta. "How to time-stamp a digital document." In *Conference on the Theory and Application of Cryptography*, pp. 437-455. Springer, Berlin, Heidelberg, 1990.

[9] Bayer, Dave, Stuart Haber, and W. Scott Stornetta. "Improving the efficiency and reliability of digital time-stamping." In *Sequences Ii*, pp. 329-334. Springer, New York, NY, 1993.

[10] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." *Decentralized Business Review* (2008): 21260.

[11] Sheldon, R. "A timeline and history of blockchain technology." *URL: https://www. techtarget. com/whatis/feature/A timeline-and-history-of-blockchain-technology* (2021).

[12] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." *Decentralized Business Review* (2008): 21260.

[13] Zhao, Hongguo, Yunxia Liu, Yonghao Wang, Xiaoming Wang, and Jiaxuan Li. "A blockchain-based data hiding method for data protection in digital video." In *International Conference on Smart Blockchain*, pp. 99-110. Springer, Cham, 2018.

[14] Nelaturu, Keerthi, Han Du, and Duc-Phong Le. "A Review of Blockchain in Fintech:Taxonomy, Challenges, and Future Directions." *Cryptography* 6, no. 2 (2022): 18.

[15] Mohsin, Ali H., A. A. Zaidan, B. B. Zaidan, K. I. Mohammed, Osamah Shihab Albahri, Ahmed Shihab Albahri, and M. A. Alsalem. "PSO–Blockchain-based image steganography: towards a new method to secure updating and sharing COVID-19 data in decentralised hospitals intelligence architecture." *Multimedia tools and applications* 80, no. 9 (2021): 14137-14161.

[16] Elisa, Noe, Longzhi Yang, Fei Chao, and Yi Cao. "A framework of blockchain-based secure and privacy-preserving E-government system." *Wireless networks* (2018): 1-11.

[17] Liu, Si, Yunxia Liu, Cong Feng, Hongguo Zhao, and Yu Huang. "Blockchain privacy data protection method based on HEVC video steganography." In *2020 3rd International Conference on Smart BlockChain (SmartBlock)*, pp. 1-6. IEEE, 2020.

[18] Sarkar, Proton, Sudipta Kumar Ghosal, and Madhulina Sarkar. "Stego-chain: A framework to mine encoded stego-block in a decentralized network." *Journal of King Saud University- Computer and Information Sciences* (2020).

[19] Alizadeh, Morteza, Karl Andersson, and Olov Schelén. "DHT-and blockchain-based

smart identification for video conferencing." *Blockchain: Research and Applications* 3, no. 2 (2022): 100066.

[20] Liu, Yiming, F. Richard Yu, Xi Li, Hong Ji, and Victor CM Leung. "Blockchain and machine learning for communications and networking systems." IEEE Communications Surveys & Tutorials 22, no. 2 (2020): 1392-1431.

[21] L. Samuel, "Some studies in machine learning using the game of checkers," IBM Journal of research and development, vol. 44, no. 1.2, pp. 210–229, 1959.

[22] Stoica, Ion, Robert Morris, David Liben-Nowell, David R. Karger, M. Frans Kaashoek, Frank Dabek, and Hari Balakrishnan. "Chord: a scalable peer-to-peer lookup protocol for internet applications. IEEE/ACM Transactions on networking 11, no. 1 (2003): 17-32.

[23] Alizadeh, Morteza, Karl Andersson, and Olov Schelén. "Efficient decentralized data storage based on public blockchain and IPFS." In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1-8. IEEE, 2020.

Links:

[h1] https://www.ibm.com/in-en/topics/what-is-blockchain#:~:text=B

lockchain%20overview,patents%2C%20copyrights%2C%20branding).

[h2] [9] R. Sharma **Bit gold** Investopedia (2021) Available online:

https://www.investopedia.com/terms/b/bit-gold.asp, Accessed 24th Oct 2021