January 22-24 2019

Japan Hokkaido

ACENS

Asian Conference on
Engineering and Natural Sciences

ISFAS

International Symposium on
Fundamental and Applied Sciences

ACENS-0297
A Combination of Convolutional Code, Distribution Cloud Storage, and Blockchain for Securing Electronic Health Records

Tanagrit Chansaeng[a], Chaiyaporn Khemapatapan[b], Pita Jarupunphol[c]
[a b] College of Innovative Technology and Engineering, Dhurakij Pundit University, Thailand
E-mail: 587191070001@dpu.ac.th
[c] Department of Digital Technology, Phuket Rajabhat University, Thailand
E-mail: p.jarupunphol@pkru.ac.th

Abstract

Although the computer network infrastructure has been designed to facilitate interoperability services for the electronic health records (EHRs) exchange, the exchange of EHRs has long been a persistent problem. General availability of data may be critical and support health research and other demographic development activities in other areas. Nevertheless, using the same data may lead to ethical issues since privacy controls are still varied in different countries. Defensibility and protection towards patient privacy against general availability are underlying principles. Blockchain technology has been acknowledged and proposed as a promising solution that is capable of maintaining accuracy and establishing credibility in data security. The blockchain provides tamper-evident and tamper-resistant digital ledgers implemented in a distributed fashion and usually without a central authority. It enables a community of users to record transactions in a shared ledger within that community, such that no transaction can be transformed once published under the blockchain network normal operation. In this article, we propose an alternative model for securing EHRs using a combination of three prominent techniques, including convolutional code, distribution cloud storage, and blockchain. This model utilised a computer tomographic (CT) method to scan pictures in the experiment. A picture was transformed from RGB to binary, encoded with convolutional code, and separated to different pieces. Essential blockchain requirements, including hash function, construct block, proof of work, and link chain were indispensable in the experiment. Poly2trellis function in MATLAB is a trellis diagram was also conducted for the decoding purpose. The results represented that the CT scan picture can be reconstructed by the code rate at 2/3 and their blocks were stored in different public cloud storages.

Keywords: blockchain, convolutional coding, electronic health recodes (EHRs), cloud storage

1. Background

The collaboration of health information has raised several issues since privacy and accessibility of data are usually contradicted with each other [2]. For instance, providing reasonable access to

electronic health records (EHRs), maintaining the privacy of personal information, disclosing information, and avoiding inappropriate use of EHRs.

A blockchain is a digital record of transactions in which individual records or 'blocks' are linked together in a single list or 'chain'. Each transaction newly added to a blockchain will always be validated by multiple computers on the Internet. These computers are configured to monitor specific types of blockchain transactions in the form of a peer-to-peer network. They cooperate together to ensure each transaction is valid before it will be added to the blockchain [7]. This decentralized network of computers ensures a single computer cannot add invalid blocks to the chain [4]. When a new block is added to a blockchain, it will be linked to the previous block using a cryptographic hash generated from the contents of the previous block. This process is to ensure that the chain will never be broken and each block will be permanently recorded. It is difficult to alter previous blockchain transactions because all the subsequent blocks must also be altered [5][6]. The blockchain technology has been increasingly applied for cryptocurrencies, such as Bitcoin, XRP, and Ethereum [3].

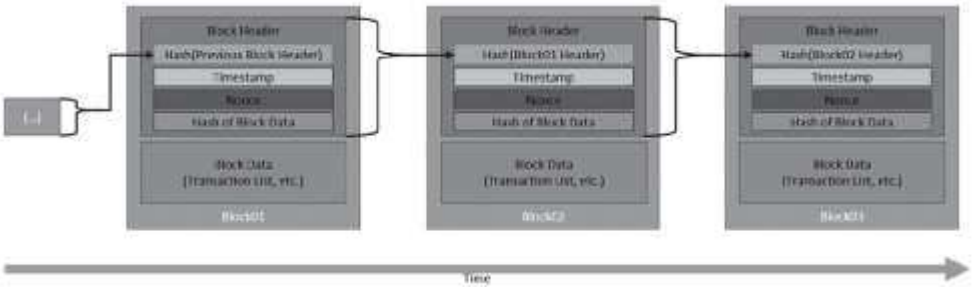


Fig. 1: A Generic Chain of Blocks [4]

This article introduced a promising technique based on a combination of convolutional code, distribution cloud storage, and blockchain for securing EHRs. We start from exploring the error correcting code for protecting patient data based on a computer tomography (CT) scan picture. After that, convolutional coding and viterbi convolutional decoding are applied to encode the picture. Primarily, the file pieces stored on the third party public free cloud storage (Google Drive, Dropbox and OneDrive) will be linked and protected using the blockchain.

Related Works

Azaria et al. [25] provided an evidence of the blockchain application as a mediator for health information. The prototype, namely MedRec, enabled medical researchers and healthcare stakeholders to 'mine' the network and obtain a mining reward for publishing syndicated and anonymous medical information. The authors asserted that a secure peer-to-peer network can be established by providing large data to empower researchers while attracting patients and providers. However, the MedRec was specifically designed to investigate medical records and

should to be expanded to generate and represent more critical health information and complex situations [8]. The Cyph MD, introduced by an Australian startup, is another prominent example of the blockchain application in health records management based on Ethereum framework [9].

GemOS is an operating system (OS) that claims to be driven by blockchain security features. There are two general ledger accounts, one for the Supply Chain and the other for the EHRs. [10]. However, a descriptive information of GemOS implementation in practice was scarce.

2. Methods

A proposed model utilises a convolutional code, which is a type of error-correcting code, to generate parity symbols via the sliding application of a Boolean polynomial function in a form of data stream. In the meantime, blockchain is used for file distribution in a free cloud environment.

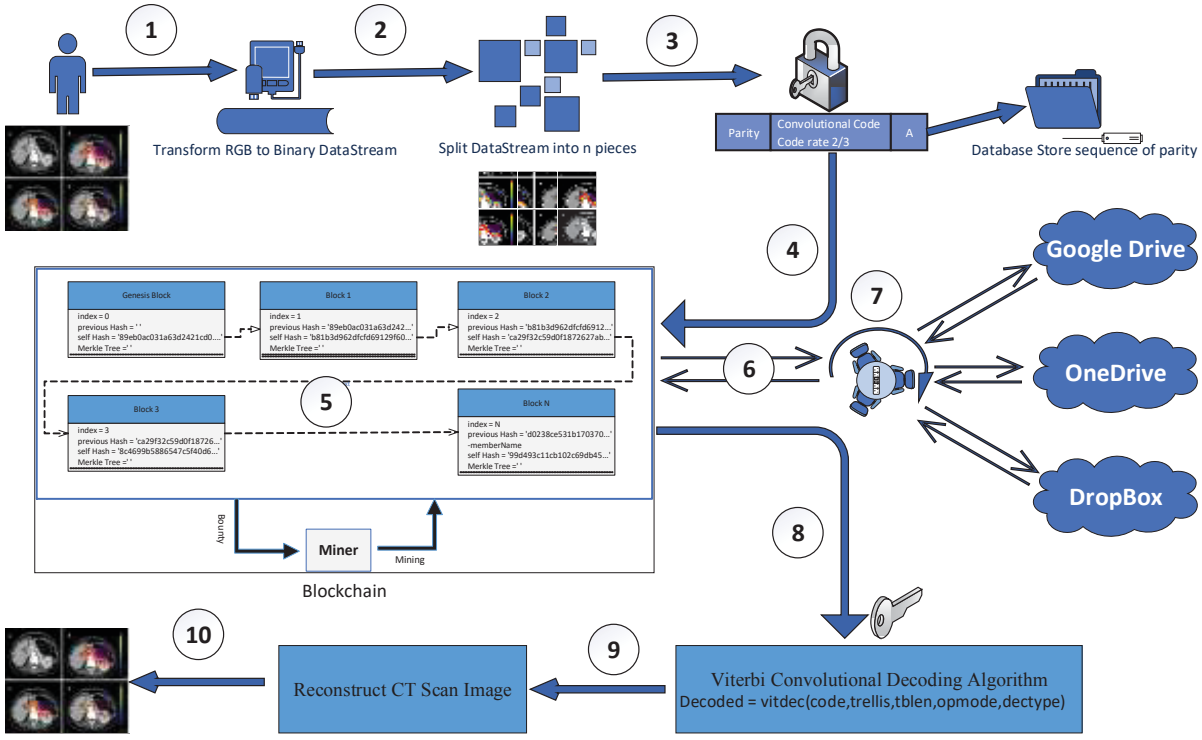


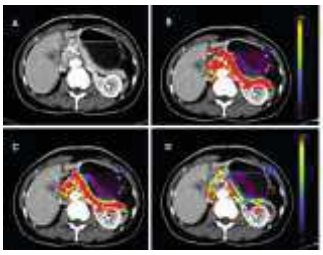
Figure 2: A proposed model

1. Data Preparation

A computer tomography (CT) is used to scan pictures from the Internet. Firstly, we transform a picture in 3 dimensions, including heigh (X), width (Y), and length (Z) to binary datastream using a Base64 encoder.

In fact, the RGB color scheme is triple-valued and allows us to measure each digit independently. All the colors are represented by the system in the form of cubes, as shown in Figure 2 [12]. Please note that the RGB color cube is black, white, red, green and blue, and

primer 3 is yellow, blue and magenta. The diagonal from the black angle to the white corner represents all grayscale. [11] Other positions within the cube correspond to other colors that can be displayed. Figure 3 represent split RGB color cube split that will print the contents of the RGB image stored in an array named blueChannel, redchannel, and greenchannel.



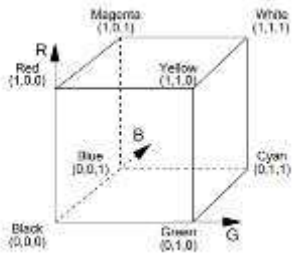


Figure 3: RGB color cube [11]

splitrgbhealth.m

```
clear all;
rgbImage = imread('InputImage3.jpg');
redChannel = rgbImage(:,:,1); % Red channel
greenChannel = rgbImage(:,:,2); % Green channel
blueChannel = rgbImage(:,:,3); % Blue channel
allBlack = zeros(size(rgbImage, 1), size(rgbImage, 2), 'uint8');
just_red = cat(3, redChannel, allBlack, allBlack);
just_green = cat(3, allBlack, greenChannel, allBlack);
just_blue = cat(3, allBlack, allBlack, blueChannel);
recombinedRGBImage = cat(3, redChannel, greenChannel, blueChannel);
subplot(3, 3, 2);
imshow(rgbImage);
fontSize = 20;
title('Original RGB Image', 'FontSize', fontSize)
subplot(3, 3, 4);
imshow(just_red);
title('Red Channel in Red', 'FontSize', fontSize)
subplot(3, 3, 5);
imshow(just_green)
title('Green Channel in Green', 'FontSize', fontSize)
subplot(3, 3, 6);
imshow(just_blue);
title('Blue Channel in Blue', 'FontSize', fontSize)
subplot(3, 3, 8);
imshow(recombinedRGBImage);
title('Recombined to Form Original RGB Image Again', 'FontSize', fontSize)
set(gcf, 'Units', 'Normalized', 'OuterPosition', [0, 0, 1, 1]);
set(gcf, 'Name', 'Demo by ImageAnalyst', 'NumberTitle', 'Off')
```

Figure 4: MATLAB Source code splitrgbhealth.m

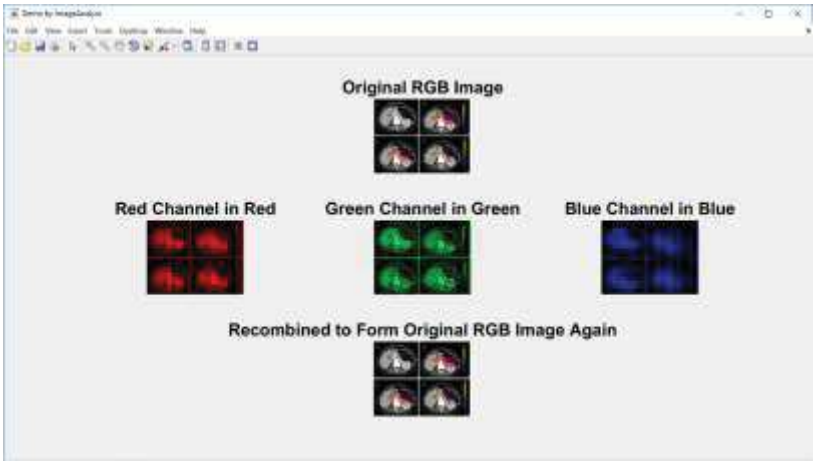


Figure 5: Result from split picture to RGB channel

2. **Convolutional Encoder**

The convolutional code is a linear combination of a valid code in bit sequences. Therefore, the distance properties of the code can be only specified by the weights of non-zero sequences. Convolutional codes are defined in number of output bits *(n)*, number of input bits *(k)*, and number of memory registers *(m)* where $n > k$ [13]. The quantity *k/n* is recognised as a code rate, which is a measure of the code efficiency. Generally, *k* and *n* parameters range from 1 to 8 and *m* ranges from 2 to 10. The code rate is varied from 1/8 to 7/8, but not applicable for deep space applications where code rates as low as 1/100 or even longer. In some cases, convolutional code specifies *n, k,* and *L* parameters. The quantity *L* is the constraint length of the code [13] [14]

Constraint Length, L = k (m-1) (1)

The constraint length *L* implies the number of bits in the encoder memory affecting the generation of the n output bits. The constraint length *L* is also referred by the capital letter *K*, which is different from the lower-case *k* that represents the number of input bits. *K* is also defined as equal to the products of *k* and *m*. In many cases, the codes are specified by *r* and *K*, where *r* = the code rate *k/n* and *K* is the constraint length [15].

Generator Polynomials: The generator polynomial specifies the connections between the shift registers and the modulo-2 adders. It is defined by

$$g^{(i)}(D) = g_0{}^{(i)} + g_1{}^{(i)}(D) + g_2{}^{(i)}(D^2) + \cdots \ldots + g_M{}^{(i)}(D^M) \quad (2)$$

where, D = unit delay variable and M = number of shift registers [13] [16]

In this paper, we used 2/3 convolutional encoder to encode the RGB channel. According to a transform instruction, we used picture size width * height pixel. Thus, all of bit is (width * height * 3) bit we call message as in (3). After that, the message is encoded by a convolutional coding as in (5) (6) and the encoded message will be lengthened according to the code rate as in (4) in

which the encoded message is divided into *n* pieces. Parity bits are added in the header of each piece of the devided picece of encoded message for 1 bit to identify a sequence of a message used for the picture reconstruction in the decode section.

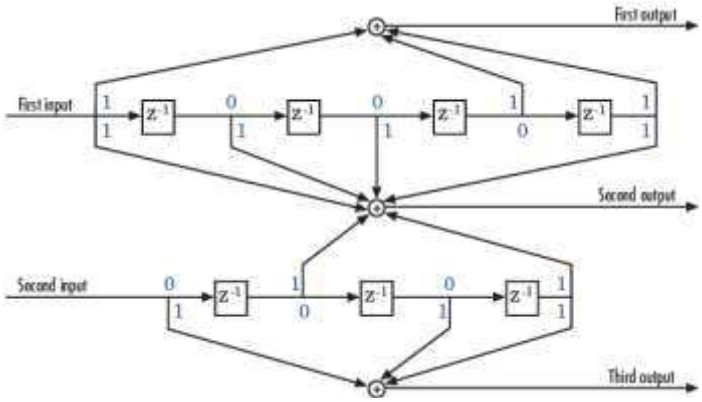


Figure 6: Schematic Rate 2/3 feedforward encoder [17]

$$message = \begin{Bmatrix} R * width * height + \\ G * 256 + \\ B \end{Bmatrix} \quad \text{bit} \qquad (3)$$

$$encoded\ message\ lenge = \frac{message*3}{2} \quad \text{bit} \qquad (4)$$

$$trel = poly2trellis([5\ 4], [23\ 35\ 0; 0\ 5\ 13]); \qquad (5)$$

$$encoded\ message = convenc(msg, trel); \qquad (6)$$

$$decoded = vitdec(code, trellis, tblen, opmode, dectype) \qquad (7)$$

3. Blockchain

The technique was designed to follow blockchain-based architected as in Figure 7. When a new block is added to a blockchain, it is linked to the previous block using a cryptographic hash generated from the contents of the previous block. This can ensure that the chain will never be broken and each block is permanently recorded. It is also difficult to alter past transactions in blockchain since all the subsequent blocks must be initially altered.

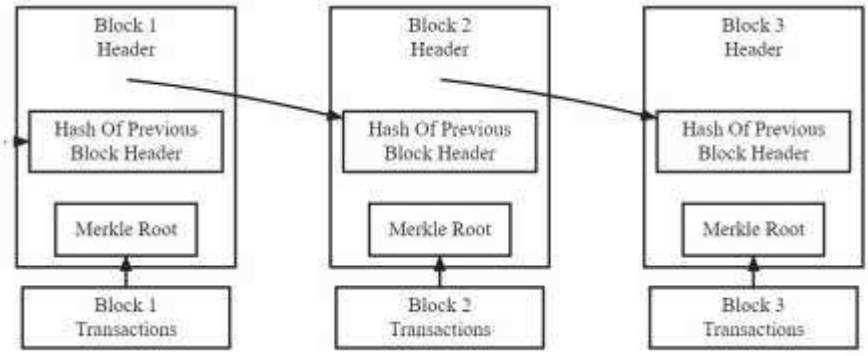


Figure 7: blockchain-based architected

Our Blockchain Prototype Procedure

A. A piece of encoded message is a transaction in term of a blockchain. A genesis block is the first block of a chain is generated and a blank block means there is no transactional data.
B. SHA-256, a cryptographic hash algorithm that converts an input to an output with a fixed size, will be used for hashing transaction with previous hash of all block except the genesis.
C. A block that has been created will be published by the miners as a new block with Blockchain Consensus Protocol (BCP), selecting a Proof-of-Work (PoW) and maintaining copies of the ledger.
D. Each block will be randomly uploaded and stored in the distributed free cloud storage such as Google Drive, OneDrive and Dropbox. Then, round-robin (RR) is deployed by the network schedulers. Time slices are assigned to each process in equal portions and in circular order, handling all processes without priority. The first process is Google Drive, second is OneDrive, and third is Dropbox.

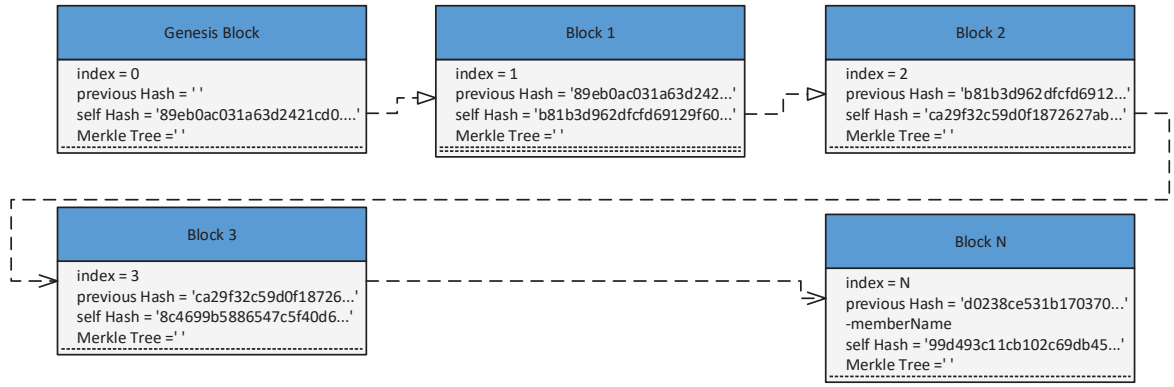


Figure 8: Our Blockchain Prototype

A Merkle tree, introduced in 1979 by Ralph Merkle, allows all computers on a network to verify individual records without having to review and compare versions of the entire database. By using cryptography that reveals an individual record, all the other records in the database are also assured with integrity protection. Merkle trees have been useful for distributed networks where different computers keep copies of the same database or ledger [18], [19].

Blockchain Consensus Protocol. We select Proof of Work (PoW) Consensus Model because it is a simple consensus model and easy to implement. In the proof of work (PoW) model, a user publishes the next block by being the first to solve a computationally intensive puzzle. The solution to this puzzle is the “proof” they have performed work. The puzzle is designed such that solving the puzzle is difficult but checking the solution validity is effortlessly. This enables all other full nodes to easily validate any proposed next blocks, and any proposed block that do not

satisfy the puzzle would be rejected. A common puzzle method is to require that the hash digest of a block header be less than a target value. Publishing nodes make many small changes to their block header (e.g., changing the nonce) trying to find a hash digest that meets the requirement. For each attempt, the publishing node must compute the hash for the entire block header. Hashing the block header many times becomes a computationally intensive process. The target value may be modified over time to adjust the difficulty (up or down) to influence how often blocks are being published. [4]

3. Results

In this section, we experimented for proof of concept for our model using MATLAB. In fact, the double process on our blockchain prototype procedure was conducted. The first time was experimented in a plane of data (encoded message) to test the model. Then, noise and burst error were randomly added to a piece of an encoded message before creating a block in the blockchain and stored in public cloud storages such as google drive, OneDrive, and Dropbox.

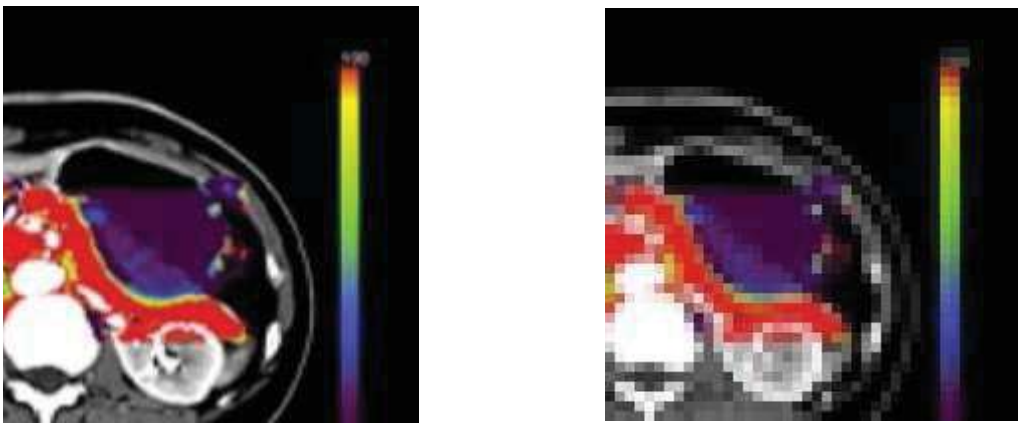


Figure 9: The piece of picture was added noise. (Left: original and Right: added noise)

In the process of decode in this paper, we use a Viterbi Convolutional Decoding Algorithm a function in MATLAB as in (7) and consistent with as in (5).

For example, we remove some bits from RGB channel for 12 bits for input. Input: 1 0 1 1 1 0 0 0 1 1 1 1 and encode with convolutional code by code rate 2/3 as poly2trellis ([5 4],[23 35 0;0 5 13]) in MATLAB a trellis diagram will construct as Figure 10. And then we got output is 1 0 0 1 1 0 1 0 1 1 0 1 1 1 1 1 1 0 1 0 1 1 0 0 0 0 0. Finally, we decode by Function vitdec in MATLAB and we got result: 1 0 1 1 1 0 0 0 1 1 1 1

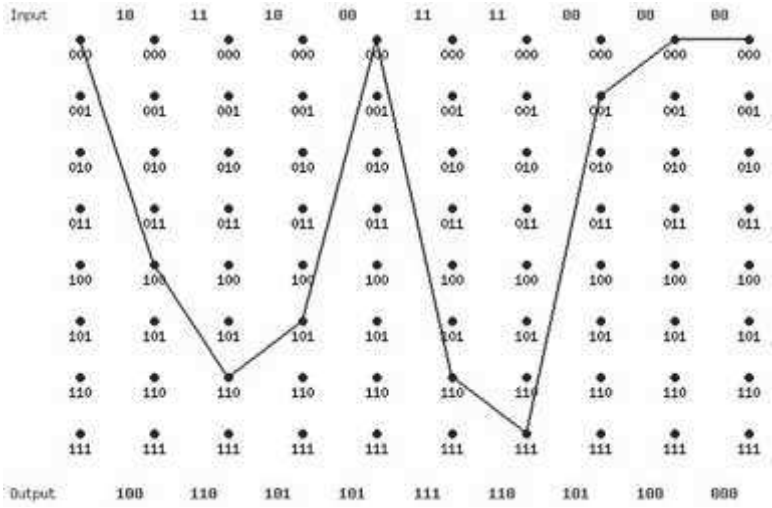


Figure 10: Trellis Diagram

The Viterbi Convolutional Decoding Algorithm was discovered and analyzed in 1967 by Viterbi [20].The Viterbi algorithm essentially performs maximum likelihood decoding. However, it also reduces the computational workload by exploiting the special structure in the code trellis. The advantage of Viterbi decoding, compared with brute-force decoding, is that the complexity of a Viterbi decoder is not a function of the number of symbols in the codeword sequence. The algorithm involves calculating a measure of similarity, or distance, between the received signal, at time t_i, and all the trellis paths entering each state at time t_i. The Viterbi algorithm removes from consideration those trellis paths that could not possibly be candidates for the maximum likelihood choice. When two paths enter the same state, the one having the best metric is chosen. This path is called the surviving path. The surviving path selection is performed for all the states [22]. The decoder continues in this way to advance deeper into the trellis, making decisions by eliminating the least likely paths. The early rejection of the unlikely paths reduces the decoding complexity. In 1969, Omura [21] demonstrated that the Viterbi algorithm is, in fact, maximum likelihood. Note that the goal of selecting the optimum path can be expressed, equivalently, as choosing the codeword with the maximum likelihood metric, or as choosing the codeword with the minimum distance metric.

Performance Analysis

The performance of convolutional codes depends on parameters such as generator polynomials, code rate and constraint length [23] [24]. The performance of a digital communication system, in terms of Bit Error rate (BER) is improved as the constraint length increases, though at the cost of increasing system complexity. As shown in Figure 11, high Eb / no error correction codes will provide a performance advantage. Figure 12 also show that at the same code rate and generator polynomials of convolutional code, it will be timely to be executed if the data is very large.

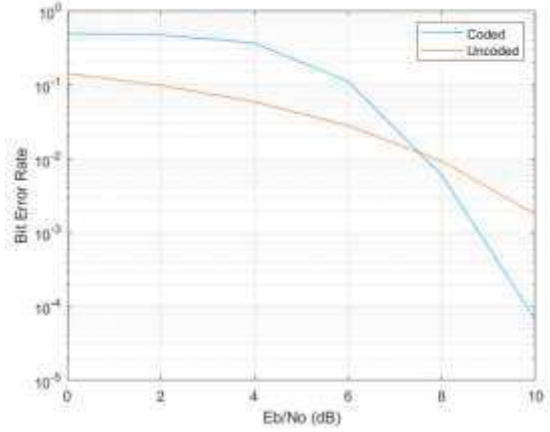


Figure 11: Plot the BER vs. Eb/No

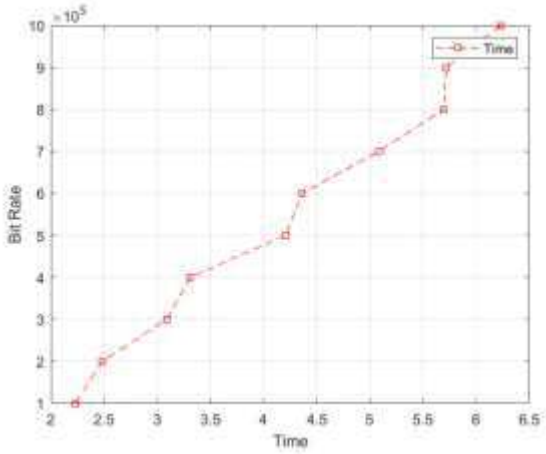


Figure 12: Plot Bit Rate vs. execute time

Conclusion

In this paper, we have proposed a novel technique for securing EHRs using a combination of convolutional code, distribution cloud storage, and blockchain technology. Our proposed technique provides a proof-of-concept and demonstrates how principles of distributed ledgers of blockchain can be provide an effective level of security for EHRs protection when CT scan images are distributed across separate public cloud storages. Nevertheless, hash values of the original picture were kept to ensure that the original picture can be obtained with integrity protection. The proposed technique can be considered as novel and secure because the blockchain fundamental concept of distribution ledger that all the nodes are required to store the same ledger. Given that the protection of patient confidential information is the first priority, an effective security mechanism must be deployed. The proposed technique separates the information into different pieces that will be encrypted using convolutional code and stored in different cloud storages. In this case, if a malicious person can obtain some pieces of the patient information, they cannot be interpreted and meaningless. Primarily, it would be extremely difficult for the malicious person to find all pieces of the information.

4. References

[1] Halamka, J. D. and Ekblaw, A. (2017). The potential for blockchain to transform electronic health records. Harvard Business Review, 3.

[2] Arlindo F. da Conceic, Flavio S. Correa da Silva, Vladimir Rocha. (2018). Eletronic Health Records using Blockchain Technology.

[3] N. Satoshi (2008). Bitcoin: A Peer-to-Peer Electronic Cash System, [Online]. Available: https://bitcoin.org/bitcoin.pdf Accessed on September 2018.

[4] Dylan Yaga, Peter Mell, Nik Roby. (2018). Blockchain Technology Overview. [Online]. Available: https://doi.org/10.6028/NIST.IR.8202 Accessed on October 2018.

[5] G. Dwyer. (2015). The economics of Bitcoin and similar private digital currencies. Journal

of Financial Stability, vol. 17, pp. 81-91, April 2015.

[6] R. Böhme, N. Christin and T. Moore. (2015). “Bitcoin: Economics, Technology, and Governance,” Journal of Economic Perspectives, vol. 29, no. 2, pp. 213-38, May 2015.

[7] M. Swan, (2015). Blockchain: Blueprint for a new economy. O'Reilly Media,Inc., 2015.

[8] Azaria, A., Ekblaw, A., Vieira, T., and Lippman, A. (2016). Medrec:Using blockchain for medical data access and permission management. In Open and Big Data (OB D), International Conference on, pages 25–30. IEEE.

[9] (2018). Sydney platform Cyph MD uses blockchain technology to facilitate data sharing in healthcare. [Online]. Available: http://www.startupdaily.net/2016/08/cyph-md-blockchain-healthcare. Accessed on October 2018.

[10] (2018). Health Blockchain technology addresses the trade-off between personalized care and operational costs by connecting the ecosystem to universal infrastructure. [Online]. Available: https://enterprise.gem.co/health/ Accessed on October 2018.

[11] (2018). Digital Image Basics. [Online]. Available: https://people.cs.clemson.edu/~dhouse/courses/405/notes/pixmaps-rgb.pdf Accessed on October 2018.

[12] Nurdan Akhan Baykan, Nihat Yılmaz and Gürsel Kansun. (2010). Case study in effects of color spaces for mineral identification. Scientific Research and Essays Vol. 5 (11), pp. 1243-1253, 4 June, 2010.

[13] S. VikramaNarasimhaReddy, Charan Kumar .K and Neelima Koppala. (2013). Design of Convolutional Codes for varying Constraint Lengths. International Journal of Engineering Trends and Technology, Vol.4, pp. 61-66, 2013.

[14] Shashank V. Maiya, Daniel J. Costello and Thomas E. Fuja. (2012). Low latency Coding; Convolutional Codes versus LDPC. IEEE Transactions on Communications, Vol. 60, No. 5, May 2012.

[15] A. J. Han Vinck, Senior member IEEE, Petr Dolezal and Young Gil Kim. (1998). Convolutional Encoder State Estimation. IEEE Transactions on Information Theory, Vol. 44, No. 4, july 1998.

[16] Sneha Bawane and V.V.Gohokar. (2014). SIMULATION OF CONVOLUTIONAL ENCODER. International Journal of Research in Engineering and Technology, Vol. 3, pp.557-561, Mar, 2014.

[17] (2018). Design a Rate-2/3 Feedforward Encoder Using MATLAB. [Online]. Available: https://www.mathworks.com/help/comm/ug/error-detection-and-correction.html#fp7855 Accessed on October 2018.

[18] J. Lee, "BIDaaS: Blockchain Based ID As a Service," in IEEE Access, vol. 6, pp. 2274-2278, 2018.

[19] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi and J. Wang, "Untangling Blockchain:

A Data Processing View of Blockchain Systems," in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 7, pp. 1366-1385, 1 July 2018.

[20] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," in IEEE Transactions on Information Theory, vol. 13, no. 2, pp. 260-269, April 1967.

[21] J. Omura, "On the Viterbi decoding algorithm," in IEEE Transactions on Information Theory, vol. 15, no. 1, pp. 177-179, January 1969.

[22] Y. Yasuda, K. Kashiki and Y. Hirata, "High-Rate Punctured Convolutional Codes for Soft Decision Viterbi Decoding," in IEEE Transactions on Communications, vol. 32, no. 3, pp. 315-319, March 1984.

[23] S. Dhaliwal, N. Singh and G. Kaur, "Performance analysis of convolutional code over different code rates and constraint length in wireless communication," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 464-468.

[24] A. Viterbi, "Convolutional Codes and Their Performance in Communication Systems," in IEEE Transactions on Communication Technology, vol. 19, no. 5, pp. 751-772, October 1971.

[25] Azaria A , Ekblaw A, Vieira T, Lippman A. MedRec: using blockchain for medical data access and permission management. International Conference on Open and Big Data (OBD) . Vienna, Austria: IEEE; 2016:25–30.