Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A PERFORMANCE OPTIMIZED REVERSIBLE CELLULAR AUTOMATA BASED SECURITY ALGORITHM FOR SECURE DATA COMMUNICATION AND STORAGE IN HEALTHCARE

SURENDRA KUMAR NANDA[1], SUNEETA MOHANTY[1], PRASANT KUMAR PATTNAIK[1]

[1]KIIT Deemed to be University, School of Computer Engineering, India

E-mail: [1]situnanda@gmail.com, suneetamohanty@gmail.com, patnaikprasant@gmail.com

ABSTRACT

Information security and protection of privacy is one of the critical tasks in any field of work. It is more critical for the field of medical science data as it involves life and critical healthcare hazards. Another aspect is the advances in technology that gives high processing capabilities to the users and they can use these high capabilities system to protect this information more effectively and efficiently. The privacy of information should be ensured not only on data stored in storage but also required during electronic communication from one place to other. The reversible cellular automata and cryptographic hash functions can be used to achieve security of data on rest and during the transition with optimal performance. We used a 128 bits encryption algorithm based on reversible cellular automation to protect information during transit and 128 bits cryptographic hash function to protect data at rest. The implementation results show these algorithms are immune to cryptanalytics and extremely difficult to break it using brute force attack. The NIST statistical results show good characteristics of this algorithm. The hardware implementation of these algorithms is simple as it uses simple logic gates and hence cost of implementation is cost-effective

Keywords: *Reversible Cellular Automata, Cryptographic Hash Function, Clinical Data Security, Healthcare*

1. INTRODUCTION

The rapid digitization in the healthcare sector produces huge volumes of data every second and it is growing with rapid speed. The structured clinical data are stored in structured digital formats like electronic health records or personal health records. With an increase of mobile users to avail of medical services online, the privacy of personal data and other clinical data is very important. This leads to a need for a security mechanism to protect these data during transmission electronically through the network and also while it is stored in a storage device.

The security mechanism is not only able to protect textual data but also able to secure medical image data. Another important parameter it should satisfy is that it will be able to optimally utilize our processing capabilities. A cryptographic encryption algorithm may ensure the privacy of data during transmission and a cryptographic hash function will be able to ensure the privacy and security of data at rest. As cellular automata inherently are a parallel computation model, an encryption algorithm and a cryptographic hash function based on cellular automation will produce optimal results in terms of security and efficiency.

The study of the literature is done in three dimensions. The first phase is focused on the literature move around the security measures in the healthcare domain. The second one is focused on the study of cellular automata-based security algorithms and cryptosystems and the final one is the study of the cryptographic hash function.

Muhammad Asif Habib [1] et al. proposed a mathematical model for access control mechanisms for medical record access. J Vora [2] et al. implemented an access control mechanism with an AT&T scheme for managing patient data. They combined this access control mechanism with ARCANA encryption for better security. Karimi Abouelmehdi [3] et al. perform an extensive survey on preserving privacy and security on big healthcare data. Deven McGraw and Kenneth D. Mandl [4] presented a privacy protection mechanism to encourage learning about the health system using digital healthcare data. Abid Haleem [5] et al. represent an overview of blockchain application in

healthcare data. The study on security mechanisms in healthcare indicates that the emphasis is on access control of clinical data and not on a complete cryptosystem to provide security of clinical data at rest and transit.

Wolfram [6] formulate a cellular automaton based pseudorandom number generator and later used it in a cellular automata-based stream cipher. Guan [7] solved various nonlinear polynomial equations which are very difficult to solve using cellular automation. Das [8] et al.in their research article presents a theoretic analysis of adaptive cellular automata. The different works of Nandi [9] et al. and Tomassini [10] et al. proposed different types of applications of cellular automata in the field of cryptography. Seredynski [11-13] et al. designed S-box using cellular automata and its use in block cipher encryption. P. Anghelescu [14] proposed a hybrid additive cellular automata bock encryption technique. D Das [15] et al. present a configurable block encryption algorithm which can be easily implemented programmatically. Naskar [16], A. Kumar [17] et al. and Ru-Jia Wang [18] presented the application of cellular automata in quantum dot CA-based Programmable logic array. Reyes [19] modelled a complex adaptive network using CA and Yangshuai Li [20] designed an integrated circuit using quantum dot which consumes less power and has better operational speed. The study on different cellular automata-based cryptosystems indicates that the performance of these algorithms may be improved by executing them in a parallel computing environment.

Merkle [21] is the first person to introduce a one-way hash function in the area of cryptography. He used it in the DES algorithm for one-way hashing. Naor and Yung [22] confirmed security results in the context of polynomial-time reducibility. So far, different hash functions based on Damgard [23] and Matyas [24] techniques have been devised.

Most of the work related to security in healthcare focuses on providing access level security and security of data within a private network only. But frequent data theft occurs in public networks and data stored in a database. Therefore, we need a cryptosystem that provides security not only at the access level but also provide security in both public and private networks including data at rest. Again, most of the literature focuses on the execution of these security systems either in a uniprocessor environment or in a distributed environment. This will provide us with an opportunity to execute it in a parallel environment using the cellular automata technique. We design a cellular automata-based cryptosystem by combining cellular automata-based hash function and reversible cellular automata-based encryption algorithm to maintain the confidentiality and integrity of data in transit and at rest. We used a standard data set "Acute lymphoblastic Leukemia image dataset" [25] for the implementation. This data set consists of Original peripheral blood smear images containing 3562 images from 89 patients.

In this work, we convert the Acute lymphoblastic Leukemia image dataset into binary data and finally used 128 bits cellular automata-based encryption algorithm and 128 bits cryptographic hash function to provide security during transmission and at rest.

The work organized in; Section II provides the background of cellular automation. We have presented our proposed work in Section III, which is to apply the 128 bits encryption algorithm and cryptographic hash function to medical data. The discussion of the findings obtained is in section IV. Lastly, Section V concludes the presented work and focuses on its future scope.

2. BACKGROUND OF CELLULAR AUTOMATION AND CRYPTOGRAPHIC HASH FUNCTION

Wolfram[6] introduced the concept of cellular automata. It is very simple to implement and hence used widely for different application areas. Many researchers used cellular automata to design cryptographic algorithms.

In this work, we used reversible cellular automata for encryption/decryption and cellular automata rules for the cryptographic hash function.

2.1 Cellular Automation

A one-dimensional cellular automaton is a discrete parallel computation model made of a finite array of 'n' cells. Each cell of the finite array communicated with other neighbour cells in a discrete amount of time 't'. The central cell 'c' changes its states St ∈ {0, 1} by implementing a local cellular automata rule and with a radius 'r1'. The neighbourhood consists of 2*r1+1 cells including the central cell 'c'. We called cellular automation one-dimensional cellular automation if the radius is 1.

Figure 1 represents the one-dimensional cellular automata using Rule 90. The change of Stage 0 to Stage 1 with periodic boundary and implementation of rule 90 depicts in figure 1.

Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

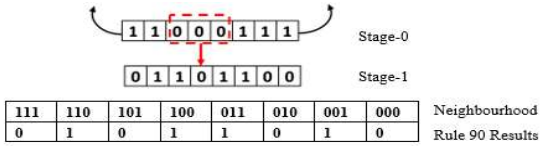

Figure 1: Cellular Automation Rule 90 with Radius=1

2.2 Reversible Cellular Automation

In reversible cellular automata, we will be able to traverse in both forward direction and backward direction using a pair of reversible rules. By applying one cellular automaton rule a configuration $Conf_i$ change its state to $Conf_{i+1}$ and by applying the complement reversible rule we will be able to traverse back from $Conf_{i+1}$ to $Conf_i$.

In cellular automation with radius 1(r1), we will be able to generate a 2^k number of rules. The value of k can be calculated by $k=2^{2*r1+1.}$

Wolfram[6] studied all rules and classified them into different classes. All these rules cannot be used for cryptographic applications. The reversible rule for a rule can be calculated using the following equation.

$R2 = 2^x - R1 - 1$, where $x=2^{2*r+1}$ and r is the radius (1)

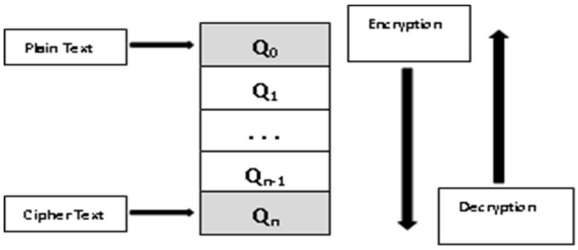

Figure 2: Forward /Reverse Movements

Figure 2 depicts the forwarding and its reverse process for the use of the cryptographic application.

2.3 Cryptographic Hash Function

Cryptographic hash functions are widely used in cryptography for one-way hashing. It accepts any number of bits as input and produces an output of 'n' bits. One way hash functions are computationally infeasible to track back to the original value and it will produce truly unique value. It is so widely used because it was very simple to implement but extremely difficult to break. A typical cryptographic hash function should fulfil the following fundamental security standards.

Given a hash value z=h(Msg1), It should be very difficult to find message Msg1' such that h(Msg') =z.

Given Msg1, it is hard to find Msg1' such that h(Msg1') = h(Msg1)

It is hard to find Msg1 and Msg1' such that h(Msg1) = h(Msg1')

3. PROPOSED METHOD

The 128 bits encryption algorithm was implemented in the HPC cluster having 64 GB of RAM and an Intel Xeon 96 cores scalable processor. The HPC cluster has Master Node with 4GB per core and two Compute Nodes with 4GB per core. OpenMP API is used 75 to achieve parallelism at the data level with both multicore and multiprocessing capabilities. We used a standard data set from Kaggle titled Acute lymphoblastic Leukemia image dataset [25].

The implementation of this method can be broken down into two parts i.e., security in transit and security at rest. Both the processes are independent of each other and combinedly provide a comprehensive security solution to clinical data.

3.1 Security in Transit

The acute lymphoblastic leukemia image dataset is first converted to binary data and then broken down into batches of 16,384 bits. Each batch consists of 128 blocks of data and each block consists of 128 bits of information. A random seed is used as the initial configuration for the implementation of the algorithm. The 128 bits random seed can be generated using any random number generator function. This algorithm requires two pairs of keys i.e., Key-1 and Key-2. Out of many reversible rule pairs only a few rules can be used for cryptographic applications. We used 153 and 195 rues as Key-1 and Key-2 respectively. The detailed process of encryption is represented in figure 3. The decryption process is exactly the reverse operation of encryption processes.

3.2 Security at Rest

The cryptographic hash function is now applied to data before storing it in the storage device. The operations of cellular automata-based hash function algorithm are broadly categorized into splitting and padding phase, Initial transformation, hashing and message digest generation.

The block diagram for the cellular automata-based hash function is represented in figure 4.

Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

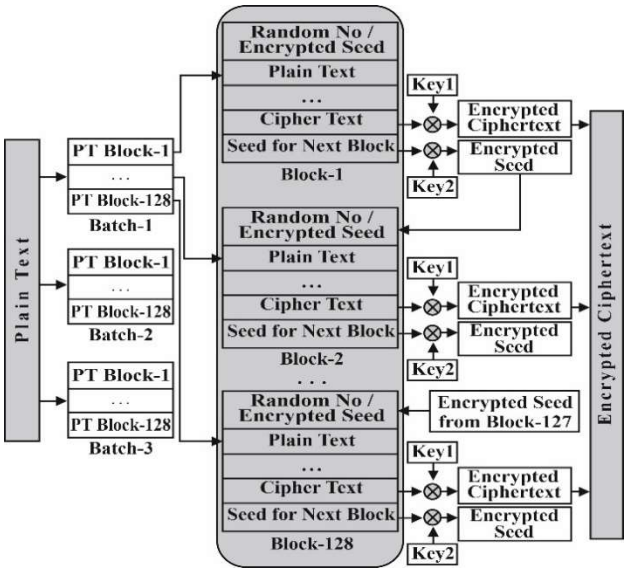

Figure 3: Detailed Process of Encryption.

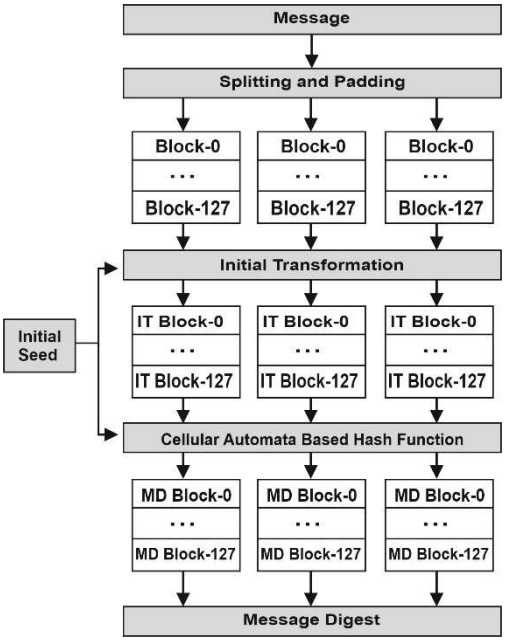

Figure 4: Block diagram of cryptographic hash function implementation

The different steps of the cellular automata-based hash function are as follows:

1. Break the data into 2KB batches
2. Break each batch into 128 blocks of 128 bits.
3. Add padding in the last batch starting with 1 followed by a continuous 0 (if required)
4. Apply the initial transformation function to each block of data.

If Block number=0

h_0=block$_0$ XOR initial seed

else

h_i=block$_i$ XOR h_{i-1}

6. Combine h_0 to h_{127} for each batch
7. Repeat 1 to 6 for all batches
8. Repeat the following activities for i from 0 to 127 (for each batch)

Find CA_{select} from hash_ruleset, where select=i MOD 7.

Calculate Message Digest$_i$ = f (CA_{select}, seed, h_i)

9. Concatenate all Message Digest$_i$ to form Message Digest for the batch.
10. Repeat Step 8 and Step 9 for each batch and concatenate all outputs to find the final message digest.

4. RESULT ANALYSIS

The 128 bits encryption algorithm was implemented in the HPC cluster having 64 GB of RAM and an Intel Xeon 96 cores scalable processor. We used a standard data set from Kaggle titled Acute lymphoblastic Leukemia image dataset. The 128 bits encryption is used for data protection during transit. We used CA rules 153 and 195 as Key-1 and Key-2 respectively. We used the 128 bits cryptographic hash function based on cellular automata before storing the data on disk. This ensures the authenticity of data stored in secondary storage.

The results of encryption, the time of encryption and decryption, and the results of the avalanche test NIST statistical test show good security and performance of these algorithms.

4.1 Performance of Execution

We run these algorithms in HPC clusters 100 times with different CPU utilization conditions and memory use patterns. These algorithms produce an average execution time of 3.533 seconds. It achieves a minimum execution time of 2.7 and a maximum of 4.7 seconds. The results of different runs are represented in figure 5.

Figure 5. Execution times of different runs.

We compare the execution time of our cellular automata-based hash function algorithm with CABHA proposed by K. Rajeshwaran and K. Anil Kumar [26] and found that the average execution speed of CABHA algorithm is 2.4 seconds in comparison to our algorithm which is 1.7 seconds. Similarly, the execution speed of our reversible cellular automata-based encryption algorithm produces 2.3 seconds in comparison to 3.8 seconds by the encryption algorithm proposed by M. Seredynski and P. Bouvry [11].

4.2 Avalanche Test

The avalanche test checks the difference in output based on a minor change of a few bits in the input. We applied the avalanche test on both algorithms by modifying 20 bits in input values. The results show both algorithms produce a substantial difference in generated output value. The results of the multiple avalanche test produce an average of 50.005 avalanche values and hence we conclude that these algorithms are statistically independent of input values. The result of the multiple avalanche test shows in figure 6.

Figure 6. Results of Avalanche Test

4.3 NIST Statistical Test

To test the pseudorandom behaviour of the algorithms we perform NIST Statistical Test. We obtain p-values from the test and found that all values are within the range of 0.001 to 0.999 and hence passed this test. The different parameters of the NIST test were found to be around 0.5 to 0.57. The details of the NIST statistical test results are shown in Table 1.

Table 1. Findings of NIST Statistical Test

Name of Statistical Test	p-value	Output Status
Frequency	0.5152	Passed
Block Frequency	0.5675	Passed
Cumulative Sums	0.5433	Passed
Runs	0.5589	Passed
Longest Run	0.5142	Passed
Fourier Transformation	0.5233	Passed

4.4 Security Analysis

A brute force attack on 128 bits encryption algorithm requires at least n^2 X 2^{128} permutations. Here 'n' is the number of cellular automata rules and the size of the random number is 128 bits. This many numbers of operations and extremely hard to perform and require huge resources. Similarly, the brute force attack on 128 bits cryptographic hash function requires 2^{128} different permutations to perform.

Again, both algorithm shows good statistical independence which makes the algorithm more secure and reliable.

5. CONCLUSION AND FUTURE WORK

The cryptosystem is implemented in an HPC cluster 100 times with different CPU utilization conditions and memory use patterns. These algorithms produce an average execution time of 3.533 seconds. The reversible cellular automata-based encryption algorithm ensures the confidentiality of data and the cellular automata-based hash function ensures the integrity and authenticity of healthcare data. The experimental results show that the 128 bits reversible cellular automata-based encryption algorithm is highly secure and shows a good degree of immunity towards brute force attacks and the throughput of this algorithm is very high as it uses the user's processing capabilities optimally. The results of the NIST statistical tool and avalanche test prove the efficiency of security. The comparison of the performance of this cryptosystem, in a parallel computing environment, with other algorithms shows better execution speed and at the same time ensures authenticity, integrity and confidentiality of healthcare data.

In future, we want to create a hardware device using this technique which will be cost-effective and can be used as a portable and pluggable device. Again, we want to upscale this algorithm to a higher block size without compromising on the throughput of the algorithm.

REFERENCES:

[1] Habib MA, Faisal CMN, Sarwar S, et al. Privacy-based medical data protection against internal security threats in heterogeneous Internet of Medical Things. *International Journal of Distributed Sensor Networks*. September 2019.

[2] J. Vora *et al*., "Ensuring Privacy and Security in E- Health Records," *2018 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 2018, pp. 1-5

[3] Abouelmehdi, K., Beni-Hessane, A. & Khaloufi, H. Big healthcare data: preserving security and privacy. *J Big Data* **5,** 1 (2018)

[4] McGraw, D., Mandl, K.D. Privacy protections to encourage use of health-relevant digital data in a learning health system. *npj Digit. Med.* 4, 2 (2021)

[5] Abid Haleem, Mohd Javaid, Ravi Pratap Singh, Rajiv Suman, Shanay Rab, Blockchain technology applications in healthcare: An overview, International Journal of Intelligent Networks, Volume 2, 2021, Pages 130-139.

[6] S. Wolfram, "Cryptography with Cellular Automata," in Lecture Notes in Computer Science, 1986, vol. 218 LNCS, pp. 429–432.

[7] P. Guan, "Cellular Automaton Public-Key Cryptosystem", Complex Systems, vol. 1, 1987, pp. 51-56.

[8] A. K. Das and P. P. Chaudhuri, "Vector space theoretic analysis of additive cellular automata and its application for pseudoexhaustive test pattern generation," IEEE Tran Comp., vol. 42, no. 3, 1993, pp. 340–352.

[9] S. Nandi, B. K. Kar, and P. P. Chaudhuri, "Theory and Applications of Cellular Automata in Cryptography," IEEE Trans. Comput., vol. 43, no.12, 1994, pp. 1346–1357.

[10] M. Tomassini and M. Perrenoud, "Cryptography with cellular automata," Appl. Soft Comput., vol. 1, no. 2, 2001, pp. 151–160.

[11] M. Seredynski and P. Bouvry, "Block cipher based on reversible cellular automata," New Gener. Comput., vol. 23, no. 3, 2005, pp. 245–258.

[12] M. Szaban and F. Seredynski, "Cellular automata-based S-Boxes vs. des S-Boxes," in Lecture Notes in Computer Science, vol. 5698, 2009, pp. 269–283.

[13] M. Szaban and F. Seredynski, "Improving quality of des S-boxes by cellular automata-based S-boxes," J. Supercomput., vol. 57, no. 2, 2011, pp. 216–226.

[14] P. Anghelescu, S. Ionita, and E. Sofron, "Block Encryption Using Hybrid Additive Cellular Automata,", 2008, pp. 132–137.

[15] D. Das and A. Ray, "A Parallel Encryption Algorithm for Block Ciphers Based on Reversible Programmable Cellular Automata," J. Comput. Sci. Eng., vol. 1, no. 1, 2010, p. 82.

[16] P. K. Naskar, S. Bhattacharyya, D. Nandy, and A. Chaudhuri, "A robust image encryption scheme using chaotic tent map and cellular automata," Nonlinear Dyn., vol. 100, no. 3, 2020, pp. 2877–2898.

[17] A. Kumar, N. S. Raghava an efficient image encryption scheme using elementary cellular automata with novel permutation box. Multimed Tools Appl 80, 21727–21750, 2021.

[18] Wang, Ru-Jia, Yi-Peng Xu, Chen She, and Mahyuddin Nasution. " A new design for programmable logic array based on QCA-based nanotechnology." Optik 253 (2022): 168581.

[19] Reyes, Leonardo, and David Laroze. "Cellular Automata for excitable media on a Complex Network: The effect of network disorder in the collective dynamics." Physica A: Statistical Mechanics and its Applications 588 (2022): 126552.

[20] Y. Li, F. Peng, G. Li and G. Xie, "A Crucial Step of Quantum-dot Cellular Automatic Placement and Routing," 2022 International Conference on Power Energy Systems and Applications (ICoPESA), 2022, pp. 96-100.

[21] R. Merkle, "One-way hash functions and DES", Advances in cryptology — CRYPTO 89, *Lecture Notes in Computer Science*,1990, vol. 435, pp. 428–446.

[22] M. Naor and M. Yung, "Universal One-way Hash Functions and their CryptographicApplications", Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing, 1989, Seattle. Washington May 15-17, pages 33-43.

[23] I. B. Damgård, "A design principle for hash functions", Advances in Cryptology — CRYPTO' 89, 1989, Proceedings Lecture Notes in ComputerScience, pp. 416–427.

[24] S. M. Matyas. C. H. Meyer. and J. Oseas "Generating strong one-way functions with cryptographic algorithm", IBM Technical Disclosure Bulletin, 1985, Vol. 27.No. 10A, pages 5658-5659.

[25] Mehrad Aria, Mustafa Ghaderzadeh, Davood Bashash, Hassan Abolghasemi, Farkhondeh Asadi, and Azamossadat Hosseini, "Acute Lymphoblastic Leukemia (ALL) image dataset." Kaggle,(2021).

[26] K. Rajeshwaran and K. Anil Kumar, "Cellular Automata Based Hashing Algorithm (CABHA) for Strong Cryptographic Hash Function," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019