Vol.65: e22210339, 2022
https://doi.org/10.1590/1678-4324-2022210339
ISSN 1678-4324 Online Edition

Article - Engineering, Technology and Techniques

Bit Incorporated Codon Positional Encoder (BICPE) Algorithm for Privacy Preservation in Vertically Partitioned Data in Cloud

Yogasini Manithurai[1]
https://orcid.org/0000-0002-7438-1171

Prathibha Badhravati Nagaraju[2]
https://orcid.org/0000-0003-2574-5730

[1]Manonmaniam Sundaranar University, Department of Computer Science and Engineering, Tirunelveli, Tamilnadu, India; [2]Manonmaniam Sundaranar University, Govindaswamy Venkataswamy Naidu College, Department of Computer Science, Thoothukudi, Tamilnadu, India.

Editor in Chief: Alexandre Rasi Aoki
Associate Editor: Raja Soosaimarian Peter Raj

Received: 01-Jun-2021; Accepted: 26-Jan-2022.

*Correspondence: yogaaarudhra@gmail.com; Tel.: +91-9659297118 (B.N.P.).

HIGHLIGHTS

- Fake transactions are inserted into database to hide item frequency.
- The input binary bits are grouped into N sets with length as four for encryption.
- Each item in a data set is encoded with A,C,G,T DNA codons.
- Encrypted Association rules are built for frequent items in a transaction.

Abstract: Current days an enormous amount of data is put away and recovered from the cloud because of its benefits like stowage size, assets pooling, and multi-occupancy. Because every administration is accessible through cyberspace, there is a lot of ambiguity about security and protection at various levels. This work proposed a privacy-preserving technique by applying Deoxyribonucleic Acid (DNA) coding-based novel Bit Incorporated Codon Positional Encoder (BICPE) algorithm for vertically divided cloud information. The transactional items are scrambled with a replacement of the DNA codon. Fictitious exchanges are appended unilaterally into the private database before encryption by the data owner to alleviate frequency analysis attacks of the items. The proposed BICPE algorithm encrypts the transactional data into A, C, G, T codons to upgrade the safety of the transactional items by replacement, operation, and intricacy. The concepts of binary coding and random values are utilized in DNA sequencing to encode the data. In this method, both the sender and the receiver will share a common DNA sequence for encryption and decryption process. The performance of the proposed technique is compared with the traditional Homomorphic Encryption (HE) in terms of support threshold values (Ts) for different items in a transaction. The exploratory outcomes demonstrate that the BICPE algorithm outperforms HE as far as ciphertext size, execution time, and privacy. Hence, the newly proposed technique is more efficient and offers better performance in terms of privacy.

Keywords: Frequent Itemset; Encryption; Eclat; Apriori; FP-growth; Association Rue Mining; DNA cryptography.

INTRODUCTION

Due to the increase in knowledge and computational control, everything is becoming digitalized. Private information is being overloaded on computers. This information is sent to different websites for appropriate applications. Information is accumulated as a joint database in the cloud by several data owners. Moreover, all this information is passed through insecure channels, and insecure problem like data hacking occur while transferring the information. In order to protect the data, some secure techniques should be adopted. Cryptography [9] is defined as the art of achieving safety by encoding information. Cryptology is defined as the process of decoding non-readable information into a readable one. Cryptology is defined as the process of grouping cryptanalysis and cryptography. Encryption is defined as the method of encoding plain information to cipher text. and decryption is defined as the reverse method of encryption.

Information analysis methods, namely Association Rule Mining and Frequent Itemset Mining, are used for identifying repeatedly co-occurring information [12]. Applications of information analysis techniques are, namely, health care [13], prediction [14], market basket analysis [15], bioinformatics [16] and web usage mining [17]. To safeguard the transactional information, correlations and hidden patterns are utilized. Each item in the transactional database has a Unique Transaction ID named “UTID”. Let J be a frequent item in a transaction only when $Support\ (J) \geq ST$. Support is defined as the number of items emerge often in the transaction. ST is defined as the Support Threshold defined by the data owner. $A \rightarrow B$ is an association rule where A and B are two separated item sets. These are the terms that show that whenever X takes place in a transaction, Y also takes place in the same transaction. Confidence of $A \rightarrow B$ is defined as the probability of how possible B is obtained whenever A is obtained. Mathematicians have developed a new-fangled information storage method called Deoxyribonucleic Acid (DNA) [10]. This method is based on DNA sequence, where one gram of DNA stores about 106 TB of information.

In this work, a novel DNA-based cryptography is adopted to secure the outsourced cloud data from unauthorized users with less computational cloud time with the regress of the key generation strategy from the bio-atomic properties of the DNA groupings. The proposed DNA-built cryptography framework is a procedure that utilizes the massive equivalent processing abilities of bio atomic calculation that converts small communications from hexadecimal and ASCII structures and afterward encodes and decodes the data. This method is based on DNA sequences, where one gram of DNA stores about 106 TB of information. In DNA, molecules are applied to execute Hamiltonian path problems, parallel operations, and computations. The information is communicated by utilizing genomic materials like DNA. The model DNA arrangement is AGCGTTGATCGTTGACGAGA. This system is intended for applications where information owners require a high end of security and is appropriate for information owners looking to outsource information stowage.

Information proprietors can outsource their scrambled information and mining assignment to an inquisitive but genuine cloud in a security safeguarding way. The support threshold value is fixed by the data owner to estimate the recurrent objects in the transaction. An encrypted association rule is generated for the frequent items in the cloud data base which is decrypted with the help of the proposed BICPE technique. In the proposed framework, the information is changed into a binary string. The binary input sequence is divided into an N number of sets, where each set is a composition of four bits. Then, at that point, the binary strings are converted into a DNA arrangement. For instance, the incomplete binary strings “00”., “01”, “10”, and “11” are changed into the nucleotides A, C, G and T, separately. Each item in a private data set is encoded with a DNA codon to conceal the things from the unlicensed clients. Aproiri, Eclat and FP-growth association rule mining algorithms are employed for improving the association rule with different k-values, where k denotes the items in a transaction. The performance of the proposed algorithm is being tested utilizing two distinctive datasets, namely Retail dataset and Pumsb dataset. The significant hidden methods in our answers are a productive and a safe outsourced examination plot. The experimental analysis illustrates that the proposed technique beats in a way that is one order higher than the other encryption algorithms as far as security is concerned with less execution time.

With the current world going digital, automated notifications are expanding. The security concern is significant when the exposure of information is allowed. Information must be privatized before it can be made accessible for information mining. Protection and security are both hindrances to information mining tasks. Numerous encryption methodologies like Secure Multiparty Computation (SMC), RSA public key cryptography algorithm, and k-anonymity have been projected to protect delicate information from unauthorized clients.

Different strategies like encryption and cryptography are adopted for covering, concealment, irritation, and anonymization of information. The authors [38] demonstrated that the gatherings scale is frequently elevated, and calculations result in significant computational and correspondence costs. A homomorphic encryption plot and a protected re-appropriated correlation conspire [12] were proposed for security, and several DNA cryptography techniques were adopted for security concerns. As of now, there are no conventional arrangements to address all protection issues correlated to each of the relevant circumstances. Exploration has been centered around discovering effective conventions for explicit issues, as it were. Nevertheless, information utility and data misfortune are compromises when viable. Motivated by all these facts, there is a need for privacy-preservation of outsourced cloud data. To overcome restrictions, in this research, a novel Bit Incorporated Codon Positional Encoder algorithm is proposed for the preservation of outsourced transactional data with less computational period and less leakage of data when compared to other privacy preservation algorithms. The experiments are carried out by varying the number of transactions such as 5000, 10000,15000 and 20000 to find frequent items in a transaction. From the experimental results, it is observed that the Eclat algorithm outperforms well.

The main contributions in the proposed BICPE algorithm are as follows:

1. This paper recommends an efficient privacy preserving proficient codon encryption conspire to work with the secure outsourced calculation support values.
2. Transactional items are scramble with a replacement DNA Codon to safeguard the administrator's crude information.
3. To improve fortification, fictitious exchanges are appended unilaterally in scrambled trade.
4. The cloud safely calculate supports and compare supports with a Threshold value Ts more resourcefully.
5. The proposed technique is compared with the existing encryption techniques.
6. Compared with the furthermost prevailing resolutions, the proposed BICPE technique accomplishes a higher security level for vertically partitioned information.
7. Although the proposed methodology is intended for the information mining arrangements illustrated in this paper. It has latent application in other secure calculation settings, like data aggregation.

There is a lot of works available related to privacy preservation of both horizontal and vertical fragmented cloud data. The techniques applied for privacy preservation with their advantages and disadvantages are listed out. [11] has employed the Modified Decrease Support of RHS item of Rule Clusters (MDSRRC) algorithm to conceal multiple R.H.S items in a distributed database. Sensitive information is confined to the cloud by providing a clean database from which the data undergoes private after the mining process. This method conceals the association rule in a database. Nomura and coauthors [27] Proposed a protected association rule mining system for vertically partitioned information. The proposed system empowers adaptable data sharing by utilizing private-set intersections. This technique utilizes scalar products. Information proprietors can share their patterns with fewer correspondence and computation costs. Domadiya and coauthors [18] has employed Privacy Preserving Distributed Association Rule Mining (PPDARM) for horizontally partitioned information on patient medical data. This method leads to perfect association results and guards the Electronic Health Record (EHR) of patients. Kalia and coauthors [19] proposed a novel Randomized Encoding (RE) strategy, where encoding is accomplished with the addition of random noise from a known dispersion to the first information. To accomplish stability between protection and information utility, the dataset traits are initially ordered into delicate and semi-identifiers. The outcome examination of the trail recommends that this technique conserves protection while sufficiently keeping up with information utility and isn't appropriate for the large dataset. [28] hypothesized a resourceful and confidentiality-conserving medical Primary diagnosis framework (CAMPS) to guarantee asylum to clients' clinical information and medical services. Exploratory outcomes show that in the current state this strategy it is essential to reduced computation and correspondence overhead.

Patel and coauthors [20]. Anticipated Hybrid k-anonymity procedure to shield the person's very own data with the least loss of data. With this technique, the unique information is altered using the randomization method and then anonymization is applied to the revised information, which can afford improved accuracy with the least loss of information. This methodology will upgrade the security of touchy data from the digital assault of the information digger. Chon and coauthors [21]. Proposed BIGMiner, a fast and scalable MapReduce-based frequent itemset mining strategy that creates equivalent estimated sub-information bases

known as exchanges chunks and achieves support counting solely based on exchange chunks and bitwise tasks without producing and rearranging midway information. BIGMiner accomplishes high versatility due to practically no memory issues. [23] has employed Pallier Encryption to safeguard the outsourced data. This encryption technique uses Rob Frugal encryption to safeguard the data owner's private data. Domadiya and coauthors [24] have applied collaborative mining to create a privacy-preserving association. Moreover, rule mining is used in vertically partitioned health care information. This technique is used to identify the relationship between a disease and its symptoms. Here, the privacy of the patients' information is saved. This technique also estimates the correlation between heart diseases and the food habits of the patients. Stergiou and coauthors [36] surveys Internet of Things (IOT) and Cloud Computing (CC) techniques with an emphasis on security issues. The two innovations have been associated to observe the intimate features and determine the advantages of their incorporation to boost the utilization and transmission of Big Data. Taking everything into account, the author concludes that the CC innovation progress the process of IOT as a base innovation for Big Data frameworks.

Suba and coauthors [25] have employed Selective Item sets Frequent Pattern Mining (SIFPMM). This proposed technique estimates the frequency of selective items in horizontal database. The execution time of three algorithms, namely FP-tree, SIFPMM, provides an exact result when compared with the other two algorithms. [34] To observe the communal features, present an overview of IOT and Cloud Computing for refuge problems of both. They suggest effective IOT and Cloud computing safety models with dual encoding calculations to advance the safety and protection problems. The endowment of distributed computing innovation advances the function of the IOT. Patel and coauthors [26] anticipated a hybrid k-anonymity procedure to shield the person's very own data with the least loss of data. With this technique, the unique information is altered using the randomization method and then, anonymization is applied on revised information which can afford improved accurateness with the least loss of information. This methodology will upgrade the security of touchy data from the digital assault of the information digger. [42] The Secured MapReduce (SMR) layer model is proposed to provide security and protection between Hadoop Distributed File System and MR Layer (MapReduce). The main advantage of this work is that it endorses information sharing for information mining. This model generates a protection and security assurance and information utility with less running time, CPU usage, memory utilization, and information forfeiture. Yin and coauthors [43] proposed a novel medical encryption procedure dependent on enhancing elliptic curve cryptography by merging it with homomorphic encryption. The trial demonstrates that this new calculation not only has a high encryption impact, high refuge, and a large key size, but it also has had a high affectability to preliminary value and against assault capacity. Furthermore, the calculation has solid strength for attacking algebraic assaults and comprehensive attacks.

There are some that look into DNA-based cryptographic methods. Subashanthini and coauthors [1] projected a tri-stage picture encoding system, which consolidates Integer Wavelet Transform (IWT), turbulent maps, and DNA encoding regulation. This strategy gives an additional degree of safety to pictures moved through different internet business application on distributed stowage. Pavithran and coauthors [2] proposed a basic cryptosystem dependent on codon-based steganography along with the fixed mechanisms hypothesis with trio units, specific to an attribute-value pair producer, a despatcher, and a recipient. Outcomes and conversation show that the proposed plot is more resourceful and secure than the current plans. Stergiou and coauthors [35] proposes a novel framework for cloud computing incorporating the internet o things as a base situation for big data. To further develop the security problems, they build up a framework that imparts on the security of the organization. According to the author, cloud computing could provide an additional "green" and productive mist climate for supportable figuring situations.

Xie and coauthors [3] Introduced a comprehensive multi-view outsourcing structure that produces numerous indistinct information views to conserve the utility of information investigation. This procedure is applied to sum up the prefix-protecting encryption to make it appropriate for more overall information type like geo-areas or DNA sequences and secure against the implication assaults. The trial outcomes establish that the projected structure conserves both security and utility with 100% investigation exactness. Leelavathy and coauthors [4] presented an Enhanced User Data Security DNA-Framework in Cloud Computing Environment (EUDSFDNA). Through applying this encryption and decoding calculation, the client ensures that the information is stowed away only on protected stowage and not recovered by others. Majumdar and coauthors [29] projected a new chromosome-based encryption method enlivened by the organic qualities of DNA and the protein blend system. Multi-Criterial Dynamic (MCDM) model was adopted for accomplishing the previously stated objective. In addition, intensive security, affectability, and functionality investigation illustrate that the anticipated conspire ensures high security with sturdiness.

On large data sets, Kannadasan and coauthors [5] used a DNA-based encryption technique. When a large quantity of information is to be saved by applying big data, then an encryption technique is applied. A DNA encoding table with the PHP language is used for the encryption process. Reza and coauthors [6] has presented an encryption method on DNA-based cryptography. In this method, data post-processing and the concepts of data pre-processing are analyzed. Because of its vigenere cipher and double layer of security**,** this technique appears to be very secure. Pavithran and coauthors [30] proposed a basic cryptosystem dependent on codon-based steganography along with the fixed mechanisms hypothesis with trio units, specific to an attribute-value pair producer, a despatcher, and a recipient. Outcomes and conversations show that the proposed plot is more resourceful and secure than the current plans**.** Majumdar and coauthors [7] proposed a DNA-based encryption method, trailed by a 256-bit Secure Socket Layer (SSL) to get information stowage. The 256-bit SSL offers protected associations during information transmission. A fuzzy-based Multi-Criteria Decision Making (MCDM) model has been utilized. This procedure can decide on the arrangement of appropriate stowage servers on which the information should be put away and brings about a decrease in the implementation period by keeping up the degree of safety to a further developed grade. Manjot and coauthors [31] proposed a DNA-based EX-3 code using the general population and private key-empowered square level for information security. The experimental results show that the complete space portion for hefty archives has been reduced and online information security has been enhanced with the application of the proposed system.

Xiao and coauthors [8] used Distributed Frequent Itemset Mining (DFIM) calculations. The dispersed processing model proposes MapReduce calculations, such as LG and RM. The exploratory results show that the RM calculation performs better in terms of calculation and site versatility, and to manipulate the DFIM calculation which is MapReduce dependent. Min and coauthors [32] suggested a proficient Boneh, Goh and, Nissim (BGN) category equal homomorphic encryption to address the security problem in the cloud atmosphere. The trial consequences demonstrate that this calculation accomplishes the highest acceleration, up to 5.3, for productive homomorphic encryption in a distributed computing setting. Stergiou and coauthors [33] studies on Big Data (BD) and Cloud Computing (CC) and their fundamental components, with an emphasis on the security and protection problems of the two innovations. This work associates the usefulness of BD and CC with the intend of analyzing the recurrent provisions and, furthermore, finding the advantages related to security issues of their incorporation. Stergiou and coauthors [37] projected an original incorporated federated model with the abbreviation InFeMo. This framework provides an energy efficient framework design and environment for clients, with the goal of increasing the amount of information on the board. The client would have less help up time in each method line by embracing the InFeMo framework.

Xu and coauthors [39] proposed EFSS, an Efficient and Fine-grained Similarity Search Scheme over encoded DNA information. An original Boolean search system is espoused to accomplish complex rationale requests, for example, merging “AND” and “NO” procedures on genes. Information access control is reinforced in the EFSS through a variation of the polynomial-based plan. Besides, the k-means clustering procedure is adopted to increase the productivity of performance. The security investigation and extensive trails show the superior accomplishment of EFSS contrasted with existing plans. Kachouh and coauthors [40] Proposed a protected cloud-based plan for stowing and dissecting grouped hereditary information utilizing Homomorphic Encryption (HE) using the Chinese Reminder Theorem (CRT). It is utilized to project a secure cloud-based application dedicated to genome data analysis. Distinctive execution and security assessments have demonstrated the effectiveness of this arrangement and its similarity to real-world applications. The execution time is decreased to the greatest extent while maintaining a high-security level. Hagestedt and coauthors [41] anticipated the novel Beacon framework to share chromosome demethylase information (MBeacon) framework, a protection by-plan approach. Broad investigations exhibit that, utilizing painstakingly picked boundaries, MBeacon can debase the presentation of the participation, thus surmising assault, altogether without considerably harming the analysis’ utility.

MATERIAL AND METHODS

Each data owner owns a private database, and data owners collaboratively mine their joint database’s association rules with the assistance of the cloud. Each data owner hides frequent data items by inserting fictitious transactions into his private database. Each data owner tags their private database’s transactions with a 1-degree Encrypted Realness Value (ERV). Information proprietors exchange their desirable k-values, and the highest value will be used for all private databases. Every item in a private database is encrypted with the DNA codon to hide the items from unauthorized users. The data owners decrypt the received seemingly frequent items’ ERVs to determine the real frequent items, which is encrypted with DNA codon.

The proposed method prototype is depicted in Figure1. Initially, counterfeit things are added into the scramble exchanges on their own. Both real and fictious information about the data owners are encrypted with the codons. In the mining state, the cloud mines affiliation rules for information proprietors in a protection safeguarding way. The cloud mines association rule candidates from the encoded combined data set. Due to the presence of fictitious exchanges, a few candidates are considered as “false positive”. The cloud authenticates candidates in a security-preserving way. The cloud proceeds all candidates and their encoded confirming outcomes, confirming them to the information proprietors. Finally, information proprietors unscramble the scrambled checking outcomes and affiliation rule candidates to recuperate the genuine association rules. The support threshold value is fixed by the sender to estimate the recurrent objects in the transaction. An encrypted association rule is also generated in the cloud database, which is decrypted with the help of the proposed BICPE technique.

Merits of the Proposed System:

The proposed DNA cryptography is one of the rapidly embryonic technologies. It presents a new desire to break strong procedures. This is because DNA processing proposals have more speed, negligible capacity, and power necessities. This calculation has high implanting capacity and is uncomplicated to bring into training. The alteration rate is lower. To break the secret information, assailants should have a huge load of data. Rather than past approaches, this one is more productive, dynamic, and performs better. To increase the level of confidentiality and intricacy, the result of concealing information in the cloud is being executed. Memory capacity and time productivity are likewise improved.

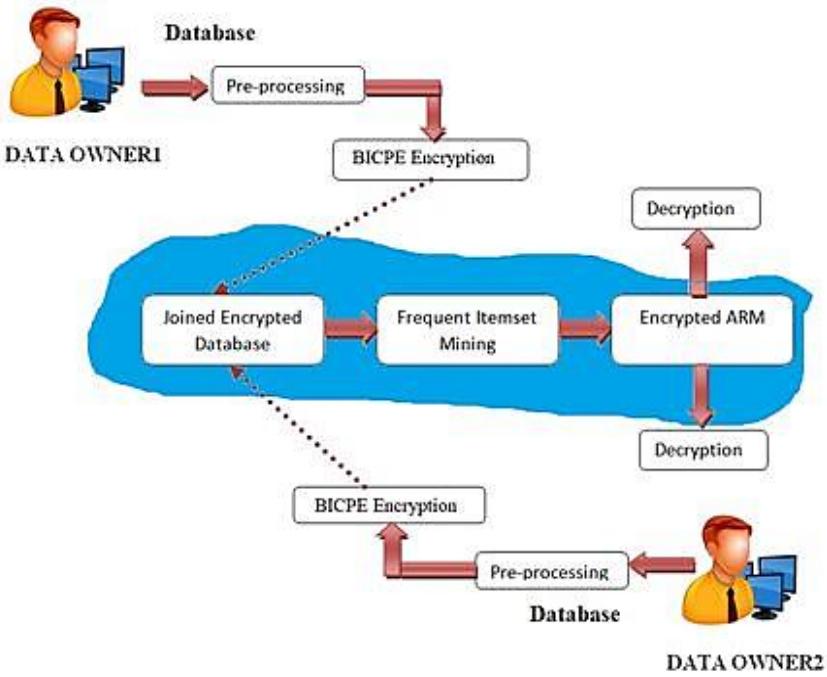

Figure1. The system model of the proposed method in cloud.

Frame Work and Design Goal

The proposed framework shown in Figure 1 consists of at least two information proprietors and a cloud. Every information proprietor has a private data set, and the information proprietors scramble their private data sets before outsourcing the encoded data set to the cloud. Information proprietors can also request that the cloud to mine affiliation handles or recurring itemset from the shared database for their benefit. The authorized, yet inquisitive cloud is entrusted with the ordering and stowing of data sets obtained from various information proprietors; the mining of affiliation leads or continuous itemsets for information proprietors; and the transmission of mining result to significant information proprietors. The cloud sincerely stores and digs information for information proprietors. Information proprietors pay for the cloud’s administrations, and they will logically pick a cloud accepted to be straightforward by a cloud supplier with a limited standing.

In the run of the mail framework arrangement for vertically partitioned data sets, information proprietors take an interest in the synergistic information mining to acquire the mining result. Every information proprietor

knows the things and the size of some other data owner's private data set and TIDs. Information proprietors ought to study as much diminutive data about data sets having a place with other information proprietors as could be expected. All the more explicitly, an information proprietor's crude exchange particulars ought not be uncovered, and the supports ought to be hidden to keep away from spillage of data about the crude information. Protection saving measures generally bring about diminished execution of information mining, and hence, any compromise and the latencies of homomorphic encryption calculation. The proposed arrangements should also shield the mining results from the cloud.

Pre-Processing

Cipher Text Substitution

The data owner's message is encrypted using the substitution cipher, which is anterior to the outsourcing method. Every item in the transaction has an equivalent substitution cipher text. Substitution cypher is an issue to frequency testing attack. A cryptographic hash function hf() is a unique kind of hash function used in cryptography. Here, hash function is applied, which maps arbitrary strings of data length of fixed output in deterministic random method.

$hf: \{0,1\}^* \rightarrow \{0,1\}^d$ string of length d where length ≥ 0

Fake Transaction

Data owners add fake items to their private messages to protect their information. To improve the protection, fake items are independently added to the encrypted transactions [22]. The data owner used a flag variable to indicate whether their transaction was a real transaction one or forgery. The flag value is set as 0 for fake transaction and 1 for actual transactions. Information proprietors embed counterfeit exchanges into their private database to ensure the information's authenticity. Fictitious transactions are added to the outsourced database so as to conceal the frequency of items. The transactional database is secured by rob frugal algorithm on the server by adding fake transaction. The security of the original transaction relies upon the number of fake transactions added to the original database, and the encrypted information contains both the original and fictitious transactions. Z fictitious transactions are embedded between every two original transactions, where Z is a random variable. The data owner labelled their transaction with an Encrypted Realness Value in the joined database and all ERV values are sent to the cloud. The flag value is set as 0 for fake transaction and 1 for actual transactions by the information proprietor.

DNA Codon Encryption

DNA Codon Table

DNA is a very lengthy molecule, completed by a long chain of nucleotides. Nucleotides are located mostly in the cell nucleus. Each nucleotide contains deoxyribose, which belongs to a phosphate group and a nitrogenous base, namely Adenine A, Cytosine C, Guanine G and Thymine T. These kinds of nitrogenous bases are situated in all nucleotide, which differentiates one nucleotide from the other. Moreover, a long chain series of nucleotides is printed as a sequence of nitrogenous bases with a related appearance in the nucleotides. The series of nitrogenous bases forms the genetic code of cells. The genetic code contains three consecutive nucleotides, which is called a codon. For example, TTT, CAG, ACT. Except for TAA, TGA and TAG, each codon of DNA employs the creation of one of 20 amino acids. TAA, TGA and TAG indicate the end of a sequence of codons. A connection of coupled amino acids and their order of arrangement is called as protein. Consequently, a series of codon in DNA molecules are coded into a specific type of protein and figure in a gene.

One gene can contain up to a thousand or more codons. Glycine has the following codons: ATA-ATC-ATT, CTA-CTC-CTG-CTT-TTA-TTG, GTA-GTC-GTG-GTT, TTC-TTT, ATG, GCA-GCC-GCG-GCT, GGA-GGC-GGG-GGT. The number of achievable codons is equal to 3^3, which means 27. The number of codons applied to create amino acids is equal to 24, which is calculated as 27 minus the three codons TAA, TGA and TAG. The number of achievable amino acids is 20, which is smaller than 24. Different codons create the same amino acid.

The Novel Bit Incorporated Codon Positional Encoder

With this model, the transactional data in the cloud is encrypted with the help of this proposed Bit Incorporated Codon Positional model. The pre-processed data is converted into binary bits for the encryption process. Then the binary input sequence is divided into N number of sets, where each set is a composition of four bits. Encryption is performed on each set separately using AGCT codons.

The Genetic code for A-Codon is ATA-ATC-ATT-ATG-ACA-ACC-ACG-ACT-AGC-AGT-AAC-AAT-AAA-AAG-AGA-AGG.

For C-Codon is CTA-CTC-CTG-CTT-TTA-TTG-CCA-CCC-CCG-CCT-CAA-CAG-CAC-CAT-CGA-CGC-CGG-CGT.

For G-Codon is GTA-GTC-GTG-GTT-GCA-GCC-GCG-GCT-GGA-GGC-GGG-GCT-GAA-GAG-GAC-GAT.

For T-Codon is TTC-TTT-TGC-TGT-TCA-TCC-TCG-TCT-TAC-TAT-TGG.

Encryption Process:

The encryption process consists of three main steps as 1. Codon Table Detection and Table Transformation, 2. Codon Table Row Selection and 3. Codon Selection and Encoding.

The input binary bits are splatted into N sets with set length as four bits. Where Set_i represents a set at the i^{th} position of N sets.

Encryption Algorithm:

T Data: Transactional Data
CT: Codon Table
Enc: Encrypted
Input: T Data, CT Contains default tables such as CT_A, CT_C, CT_G, CT_T
Output: Encrypted Data

Begin

1. **Procedure** Detection (Default CT)
2. Binary ← Bin (Data)
3. Set_i ← Split Binary Data into N sets (Where set length = 4 bits)
4. $m = mod\ (i, 4)$ for each $set\ i \in N\ sets$
5. Default $CT_{selected} \leftarrow CT_A, if\ m == 1$
6. Default $CT_{selected} \leftarrow CT_C, if\ m == 2$
7. Default $CT_{selected} \leftarrow CT_G, if\ m == 3$
8. Default $CT_{selected} \leftarrow CT_T, if\ m == 0$
9. **Procedure** Transformation (CT)
10. $CT_{selected}$ ← Default CT , $if\ first\ bit\ (Set_i\ [1]) = 0$
11. $CT_{selected}$ ← Next $CT(CT_{selected})$, $if\ first\ bit\ (Set_i[1])$ = 1
12. **Procedure** Row Selection (CT)
13. $CT_{row} \leftarrow Rand\ (Round\ trip\ (CT_{selected}))$
14. **Procedure** Selection (Codon)
15. $CT\ Codon_{ind} \leftarrow CT\ Codon_{ind} + bin_2dec\ ([2:4]) \leftarrow CT_{row}$
16. Enc data ← $[CT\ Codon_{ind}]$

End

Let us consider there is N number of sets formed from the cipher text. The original data set_i represents a set of the i^{th} position on N sets.

Detecting Codon Table:

The initial or default codon table is selected on the basis of i^{th} position. Modulo division of i with the set length (4) is used to select the codon table. The procedure to select the default codon table is described in the algorithm from step 4 to step 8.

Codon Table Transformation:

The codon table transformation is performed on the basis of first bit of Set_i based on steps 10 and 11 of the encryption algorithm.

For example, consider a Set_5 which is 1010.

The value 1010 is set to B1, B2, B3 and B4 respectively.

The value of m is 1, A codon is selected as default.

But, the bit1 value =1, the default A codon changes to C

Codon Table Row Selection:

The following example is given to illustrate step 13 in the algorithm.

Pick a random row number r_n in the range of 1 to 'n' number of rows in the selected codon table, where r_n = 4 from the range 1 to5.

The 4th row of C codon is selected as detailed below.

1. CTA, CTC, CTG, CTT, TTA, TTG,
2. CCA, CCC, CCG, CCT
3. CAA, CAG
4. CAC, CAT
5. CGA, CGC, CGG, CGT

Codon Selection:

The bits selected for consideration are Bit2-Bit4. The decimal equivalent of these three bits is used to find the Codon Position (CP) from the randomly picked row of the selected codon table. CP is estimated as shown below as per the step 15 of the algorithm.

Fetch the codon from the selected table at the position of CP by traversing from the first codon of the randomly picked row in Round Trip manner.

By considering the last three bits, the decimal value obtained is 2. (B2-0, B3-1, B4-0)

2nd position element CAT from the 4th row of the C_codon is selected for encryption as detailed below.

1. CTA, CTC, CTG, CTT, TTA, TTG
2. CCA, CCC, CCG, CCT
3. CAA, CAG
4. CAC, CAT
5. CGA, CGC, CGG, CGT

The input set_5 is encrypted as CAT.

1010 = CAT.

If the Set_5 = 1111

Consider the last three bits, the decimal value is 7. (B2-1, B3-1, B4-1)

It is seen from the above outcome, there are only two positions in the 4th row.

Counting started from the 4th row, then 5th row.

The needed 7th position is in the 1st row.

CTA codon is selected for encryption as below.

1. CTA, CTC, CTG, CTT, TTA, TTG
2. CCA, CCC, CCG, CCT
3. CAA, CAG
4. CAC, CAT
5. CGA, CGC, CGG, CGT

Decryption Process:

The decryption process consists of three main steps as 1. First Bit Formation, 2. Codon Table Row Selection and, 3. Positional Bit Formation.

Decryption Algorithm:

Enc: Encrypted
Dec: Decrypted
CT: Codon Table
Diff: Difference
Input: Enc Data
Output: Transactional Data
Begin
1.**Procedure** Formation (First Bit)
2. $m = mod\ (i, 4)$ for each $Codon_i \in CT$
3. $Dec\ first\ bit\ (Codon_i) == 0, if\ m == 1\ \&\&\ Codon_i\ \in\ CT_A$
4. $Dec\ first\ bit\ (Codon_i) == 1, if\ m == 1\ \&\&\ Codon_i \notin\ CT_A$
5. $Dec\ first\ bit\ (Codon_i) == 0, if\ m == 2\ \&\&\ Codon_i \in CT_C$
6. $Dec\ first\ bit\ (Codon_i) == 1, if\ m == 2\ \&\&\ Codon_i \notin CT_C$
7. $Dec\ first\ bit\ (Codon_i) == 0, if\ m == 3\ \&\&\ Codon_i \in CT_G$
8. $Dec\ first\ bit\ (Codon_i) == 1, if\ m == 3\ \&\&\ Codon_i \notin CT_G$
9. $Dec\ first\ bit(Codon_i) == 0, if\ m == 4\ \&\&\ Codon_i \in CT_T$
10. $Dec\ first\ bit\ \left(Codon_{i)}\right) == 1, if\ m == 4\ \&\&\ Codon_i \notin CT_T$
11. **Procedure** Row Selection (CT)
12. $CT_{row} \leftarrow Rand\ (Round\ trip\ (CT_{selected}))$
13. **Procedure** Formation (Bit 2 to Bit4).
14. $Diff =\ CT_{row} - Index\ (CT(Codon_i))$
15. $for\ j = 2\ to\ 4$
16. $bits[j] = bits[j] + dec_2bin, if\ (Diff > 0)$
17. $bits[j] = bits[j] + dec_2bin\left(CT_{row} + abs(Diff)\right), if\ (Diff < 0)$
18. $Dec\ data\ \leftarrow bits[j]$
End

First Bit Formation:

Consider N codons available in the encrypted sequence.
$Codon_i$ is in the i_{th} position of N codons. The bit 1 formation is calculated as shown in algorithm.
Consider the encrypted 5th codon CAT.
The encrypted codon CAT is not present in the first row of the A codon.
As per the above algorithm, the value of first bit = 1.

Codon Table Row Selection:

Decryption process adopt random seed method.
Select one arbitrary number r_n series of 1 till 'n' number of rows in the codon table where the codon occurs. 4^{th} row in the C_codon is selected as in step 12.
The procedure to generate ciphertext for the last three bits as shown in step 14.
In this regard, check the row and position of the input in the codon table. The 4th row has selected as per the r_n value is shown below.

1. CTA, CTC, CTG, CTT, TTA, TTG
2. CCA, CCC, CCG, CCT
3. CAA, CAG
4. CAC, CAT
5. CGA, CGC, CGG, CGT

The selected codon is CAT. The position of CAT is 2nd position of the 4th row.
Diff value is 2 is a positive value.
The binary equivalent of the resultant decimal number 2 is 010.
Last three bits (B2-B4) are considered for the calculation.
Bit 1 value = 1, remains as it is.
The decrypted cipher text of Set_5 = 1010.

The proposed decryption technique is the solid procedure for information refuge as it has fast computation time. This procedure is assigned to the cloud server for adaptability and low computational intricacy on the information proprietor side, guaranteeing speed. The time taken to decode the encode the cloud information is too long for a life period. The encoded data must be decoded without an exceptional unscrambling key. The information proprietor can decrypt the information without key generation. This strategy can save the length of the DNA succession by continually possessing the payload at nothing. Additionally, it has a more productive limit since it substitutes some of the DNA nucleotides with restricted information bits or different nucleotides depending on the confidential information being accomplished.

RESULTS

Dataset:

The trial assessment is executed for retail and the pumsb repository. They are non-proprietary repositories. The retail dataset holds 88,162 exchanges and the pumsb data set contains 49,046 transactions. The former data set is Belgain retail store and the later one is census data.

The proposed encryption technique is applied to the pumsb and retail datasets. Complexity calculations are performed using these datasets. This work also focuses on comparing the cloud execution time of three mining algorithms, namely Apriori-BICPE, Eclat-BICPE, and FPGrowth-BICPE, on vertically partitioned transactions. The computation cloud time of four different transactions with different k-values is evaluated. With the help of Python, the performance of these three algorithms is evaluated independently for the specified support threshold and confidence of items in the transaction set. The transaction sets taken for the experiment were 5000, 10000, 15000, 20000 with various k-values.

Table 1. Outsourced computation time for Pumsb dataset under 5k and 10k exchanges.

Transaction	**5k**						**10k**					
Min Support	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgro Homo	Fpgro BICPE	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgrowth Homo	Fpgrowth BICPE
0.3	4.3	3.4	5.7	4.2	6.3	5.8	5.6	4.4	6.6	5.6	7.3	6.4
0.5	3.6	2.9	4.2	3.7	5.6	4.4	4.2	3.5	5.5	4.5	6.5	5.4
0.7	2.3	2	3.5	2.8	4.4	3.6	3.5	2.7	4.7	3.5	5.4	4.8
0.9	1.3	1.1	2.4	1.7	3.4	2.7	2.4	1.4	3.4	2.9	4.2	3.5

Table 2. Outsourced computation time for Pumsb dataset under 15k and 20k exchanges.

Transaction	**15k**						**20k**					
Min Support	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgrow Homo	Fpgrow BICPE	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgrow Homo	Fpgrow BICPE
0.3	6.5	5.4	7.7	6.23	8.1	7.6	7.6	6.3	7.9	6.4	9.1	8.3
0.5	5.5	4.1	6.3	5.4	7.7	6.1	6.3	5.1	6.2	5.4	8.7	7.3
0.7	4.2	3.8	5.1	4.7	6.9	5.9	5.1	4.9	5.2	4.9	7.1	6.1
0.9	3.7	2.3	4.3	3.6	5.9	4.3	4.7	3.5	4.8	3.7	6.3	5.4

For the pumsb dataset, Table 1 compares the Eclat algorithm with homomorphic encryption, Eclat with BICPE encryption, Apriori with homomorphic encryption, Apriori with BICPE encryption, FP-growth with homomorphic encryption, and FP-growth with BICPE encryption. Here the transactions considered are 5000 and 10000. The minimum support values are 0.3, 0.5, 0.7 and 0.9 respectively. The results show that the combination with the proposed methods gives efficient results when comparing the algorithms with the homomorphic encryption combination. While considering minimum support of 0.9 with 10k transactions, the result obtained in Eclat with BICPE encryption is 1.485 sec, which is minimum cloud execution time. The proposed method has reduced cloud execution time by 29%. Among the three mining algorithms, Eclat provides the most reduced time. Mostly, Eclat algorithm is applied to all types of data set. But, Apriori is suitable only for large datasets. Here, the Eclat algorithm with the proposed BICPE has produced a reduced cloud execution time while comparing with the other combinations.

Table 2 represents the combination of the three mining algorithms and the proposed method for the Pumsb dataset. Here, the transactions considered are 15k and 20k respectively. While considering minimum support of 0.7 with 2k transactions, the result obtained in Eclat with BICPE encryption is 4.941 sec, which is the minimum cloud execution time. Next to this, Apriori with BICPE encryption has obtained is 4.951 sec. The proposed method has reduced cloud execution time by 21% while considering a combination with the homomorphic encryption algorithms. The Apriori algorithm performs an iterative approach with a candidate key generation. So, the Apriori algorithm has a longer execution time in calculating the frequent items in a transaction. When comparing with the three mining algorithms, FP-growth has a larger execution time due to the FP tree construction. For 15k and 20k transactions, the Eclat algorithm with the proposed BICPE results in reduced cloud execution time.

Table 3. Outsourced computation time for Retail dataset under 5k and 10k exchanges.

Transaction	**5k**						**10k**					
Min Support	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgrow Homo	Fpgrow BICPE	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgrow Homo	Fpgrow BICPE
0.3	5.1	4.3	6.3	5.1	7.5	6.2	6.8	5.1	7.3	6.2	8.1	7.1
0.5	4.3	3.3	5.5	4.4	6.7	5.2	5.2	4.9	6.1	5.3	7.1	6.5
0.7	3.7	2.2	4.5	3.4	5.3	6.3	4.9	4.3	5.2	4.1	6.7	5.3
0.9	2.1	1.1	3.4	2.7	4.1	7.5	3.3	3.2	4.9	3.1	5.7	4.2

Table 3 represents the comparison between the combinations of Eclat algorithm with homomorphic encryption. For retail datasets, Eclat with BICPE encryption, Apriori with homomorphic encryption, Aprioir with BICPE encryption, FP-growth with homomorphic encryption and FP-growth with BICPE encryption. Here the transactions considered are 5k and 10k. The minimum support values are 0.3, 0.5, 0.7 and 0.9 respectively. The results show that the combination with the proposed methods gives efficient results when comparing the algorithms with the homomorphic encryption combination. While considering minimum support of 0.3 with 5k transactions, the result obtained in Eclat with BICPE encryption is 4.251 sec, which is the minimum cloud execution time. But FP-Growth with homomorphic encryption has obtained 7.521 sec. The proposed method has reduced the running time by 23%.

Table 4 represents the combination of the three mining algorithms with the proposed method for retail dataset. Here, the transactions considered are 15k and 20k, respectively. While considering the minimum support value 0.5 with 10k transactions, the result obtained in Eclat with BICPE encryption is 6.284 sec, which is the minimum cloud execution time. Next to this, Apriori with BICPE encryption has obtained 7.018 sec. The proposed method has reduced the running time to 19% while considering the combination of homomorphic encryption algorithms.

Table 4. Outsourced computation time for Retail dataset under 15k and 20k exchanges**.**

Transaction	**15000**						**20000**					
Min Support	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgrow Homo	Fpgrow BICPE	Eclat Homo	Eclat BICPE	Apriori Homo	Apriori BICPE	Fpgrow Homo	Fpgrow BICPE
0.3	7.3	6.4	8.3	7.5	9.5	8.4	8.3	7.5	9.2	8.1	10.6	9.2
0.5	6.1	5.9	7.5	6.6	8.1	7.2	7.5	6.2	8.2	7.1	9.3	8.6
0.7	5.6	5.1	6.4	5.4	7.7	6.3	6.1	5.2	7.6	6.3	8.6	7.4
0.9	4.2	4.6	5.5	4.1	6.1	5.1	5.6	4.6	6.9	5.2	7.5	6.5

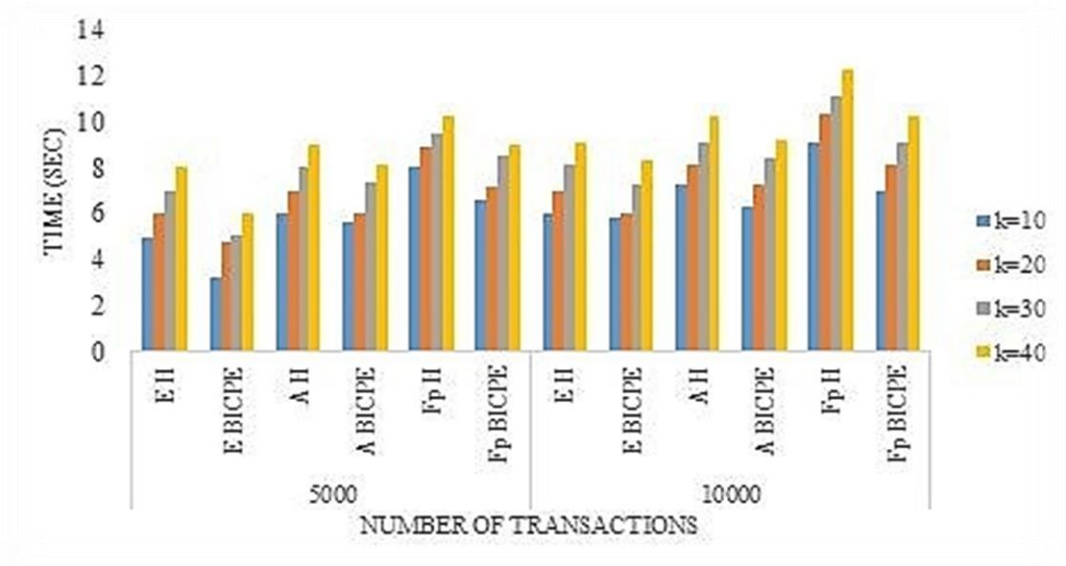

Figure 2. Cloud Execution time with different k-values for Pumsb Dataset using 5k and 10k exchanges.

In Figure 2, Like minimum support values, the running time for different k-values are evaluated. In view of k to be 30 along with 10k exchanges, outcome attained is for Eclat BICPE is 7.2 sec, with minimum cloud execution time. The outsourced computation period to pumsb repository hardly fluctuates. The support of pumsb dataset is already very high without including pretended information.

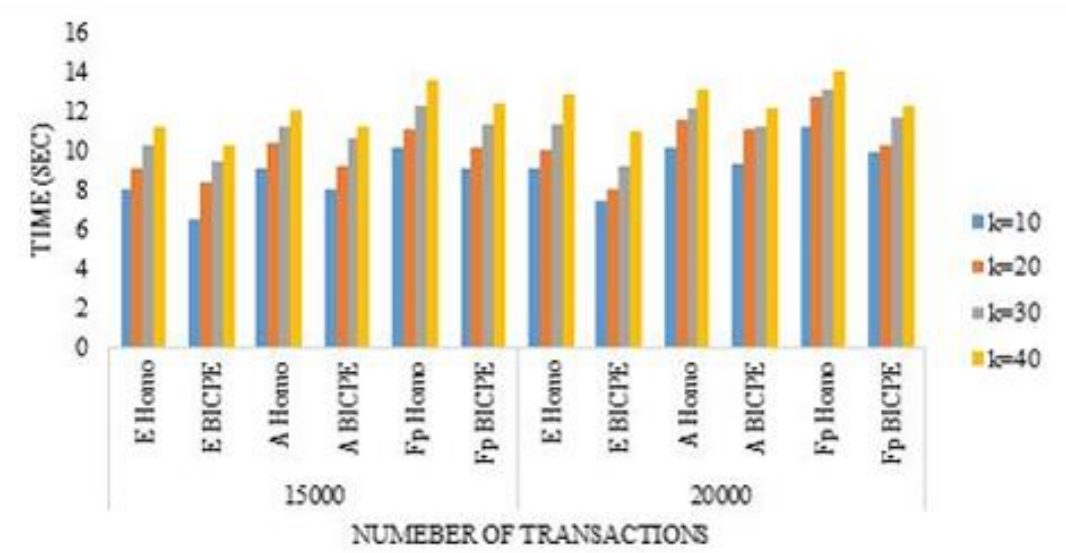

Figure 3. Cloud Execution time with different k-values for Pumsb Dataset using 15k and 20k exchanges

In Figure 3, assuming a k value of 10 with 15k transactions, the result obtained is 6.5 sec with Eclat BICPE encryption, which has a minimum cloud computation time. The cloud's run time hardly changes for pumsb dataset. The pumsb dataset is very opaque. The support of the pumsb dataset is already very high without including pretended information. Among the three mining algorithms, Eclat provides reduced execution time.

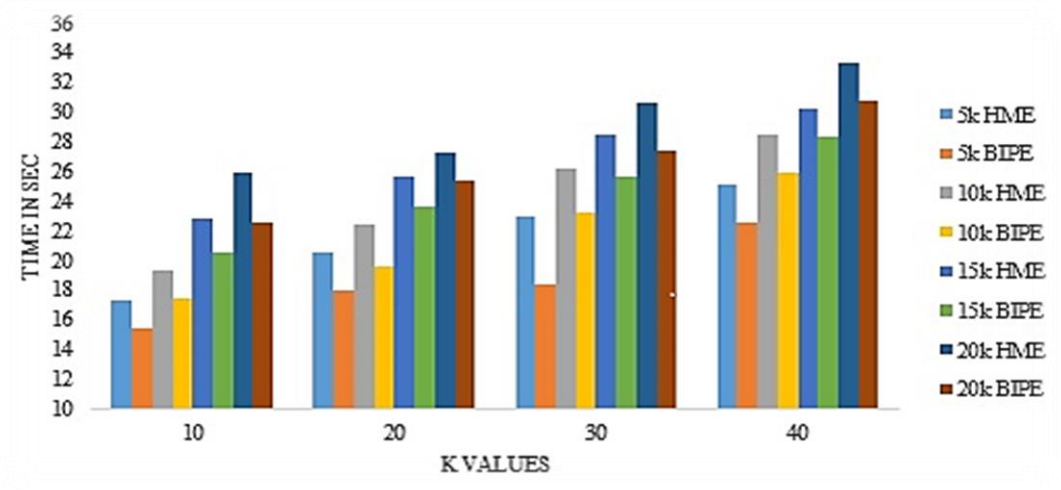

Figure 4. Data owner Execution time with different k values for pumsb dataset.

Data owner's encoded data sets are transferred to the cloud, and these data sets don't share any items. The information proprietors then, at that point, decode the encoded confirming outcomes and frequent item candidates to recuperate the genuine regular items. The information proprietor unscrambles the conventional itemsets with the reviewed supports higher than the recurrence threshold and produces affiliation rules dependent on the found continuous itemsets. To guarantee a reasonable assessment, the information proprietor has similar computer hardware and programming settings. The data owner side decrypting time for various transactions through k values for the pumsb repository is represented in Figure 4. Even though k-value increases, the Eclat algorithm with the proposed method has fewer data owner execution time.

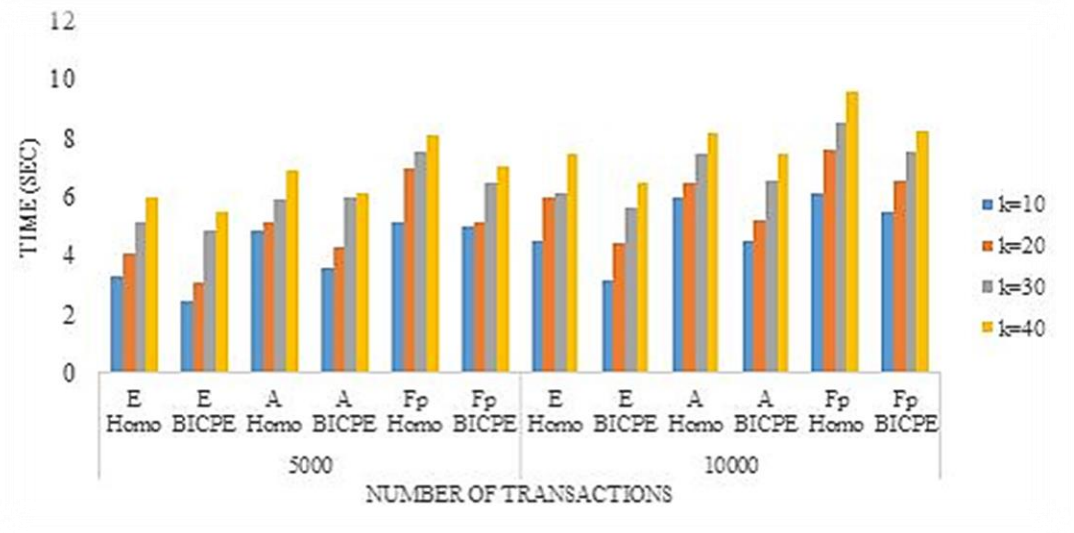

Figure 5. Cloud Execution time with different k-values for Retail Dataset using 5k and 10k exchanges.

From Figure 5 and Figure 6, it is observed that the running time changes with increasing values of k. The cloud's running time increases with k for the retail dataset. The increase in running time for retail dataset is due to the increase in pretended information. Figure 6 also represents the results of various transactions with k-values for the retail dataset. In Figure 5, while considering k-value as 40 with 5k transactions, the result obtained in Eclat with BICPE encryption is 5.458 sec, which is the minimum cloud execution time. In Figure 6, assuming k of 10 along with 20k exchanges, the result obtained is 5.476 sec along Eclat BICPE, with minimum outsourced implementation period.

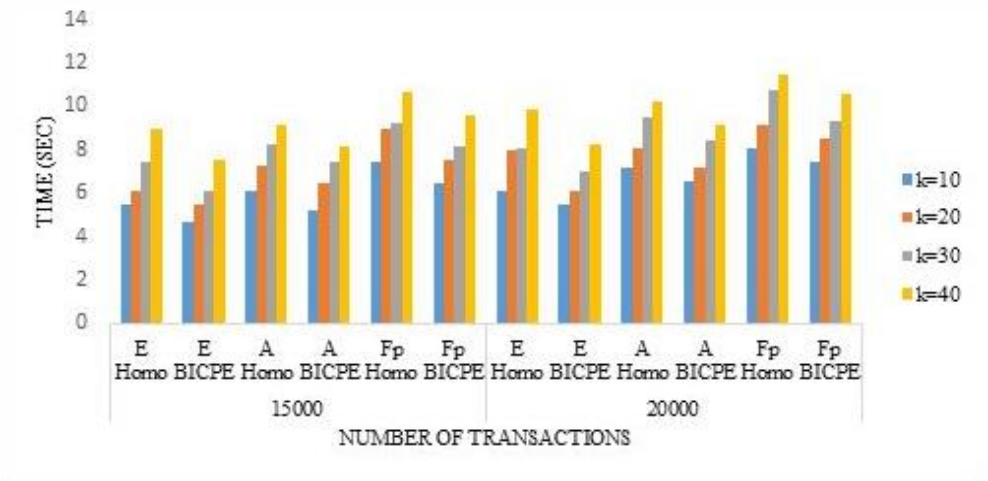

Figure 6. Cloud Execution time with different k-values for Retail Dataset using 15k and 20k exchanges.

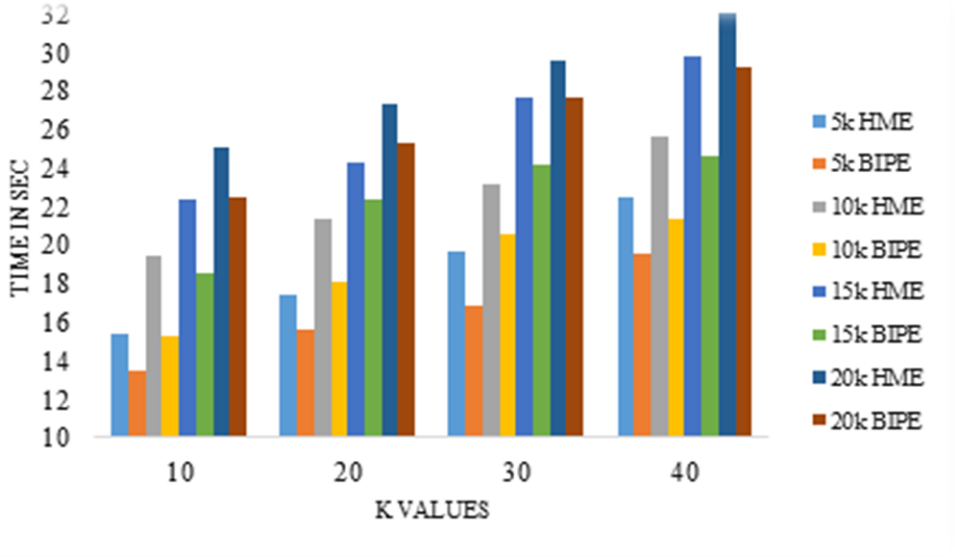

Figure 7. Data owner Execution time with different k values for Retail dataset.

The cloud yields all candidates and their encoded consequences to the information proprietors. Finally, information proprietors unscramble the scrambled checking outcomes and affiliation rule candidates to recuperate the actual affiliation rules. The information proprietor decrypts the acknowledged itemsets with the updated support higher than the recurrence threshold and produces affiliation rules dependent on frequently found itemsets. The data owner-side decrypting implementation period is represented in Figure 7 for the retail dataset. The assessment has additionally established that the answers are exceptionally effective. Consequently, the results are appropriate to be utilized by information proprietors wishing to outsource their data sets to the cloud. They require a significant degree of protection without negotiating on execution.

DISCUSSION

A novel algorithm for the preservation of vertically partitioned outsourced data is adopted for Bit Incorporated Codon Positional Encoder (BICPE). The objective is to guarantee that every item shares a similar recurrence with at least k-1 things. When compared to Homomorphic Encryption (HE), the proposed algorithm provides a high level of privacy. HE uses the symmetric key, whereas the proposed BICPE encrypts the text without key generation. This leads to less memory storage with less implementation period. The memory capacity needed for HE is 8 GB, whereas for BICPE memory needed is 6 GB. This projected method reduced the computation period by 23% for 5k and 10k exchanges and by 19% for 15k and 20k exchanges for the retail dataset. Similarly, the running period for the pumsb dataset has been reduced to 29% for 5k and 10k exchanges and 21% for 15k and 20k exchanges. The experimental assessment exhibited that this solution is exceptionally more productive than HE. Consequently, this technique is reasonable to be utilized

by proprietors wishing to outsource their data sets to the cloud yet require a significant degree of security without negotiating on performance.

Security Analysis of the Proposed Solution

In the proposed solution, the transactional items are replaced with a substitution cipher in the pre-processed state. Item frequency is hidden by adding fictitious transactions to data owners' private database to counter frequency analysis attacks. Security of the items can be achieved by adding more fictitious transactions to the transactional database. Cryptographic techniques are adopted to provide security for valuable information, such that only an authorized person can decode the information. The pre-processed data is encoded with codons in the A,C,G, and T codon tables. To prevent the attacks, the BICPE technique configures bit length for the binary values to compound the challenges of guessing the correct values. The decimal equivalent of the binary values is taken into consideration for encryption.

The round-trip random number generation method is used to select a row from the codon table where the targeted codon is selected for encryption. It increases the security level of the items in a transaction. In this method, the sender and the receiver will share a mutual DNA codon for the encryption and decryption process. It is hard to find encrypted outsourced data for unauthorized users. Compared with most existing solutions like HE, the proposed solution leaks less information about the data owner's raw data. The assessment has also demonstrated that this method is exceptionally proficient. Consequently, this solution is appropriate to be used by information proprietors wishing to outsource their database to the cloud but require a significant degree of privacy without compromising on performance.

CONCLUSION

In this paper, privacy-preserving outsourced encryption method for vertically partitioned database has been proposed. Here, the database goes through data item encryption using DNA cryptography-based algorithm known as novel BICPE algorithm. Fake transactions are included in the transaction to hide frequent items. The item's support value is compared with the Threshold Value Ts to find frequent items in a transaction. Each item in a private data set is encoded with a DNA codon to hide the items from the unlicensed clients. The proposed method obtains the same level of security as the current state of work. It is highly secured without key generation, shifting, and round-trip techniques. As the codon selection varies, it is difficult to figure out which codon is to be selected. The codon depends on the random number generator.

The position of the random number generator generates a row and the targeted codon is selected from that row. As the row has a random number, the codon to be selected depends on the content. It increases the security level of the items in a transaction. It is hard to find encrypted outsourced data for unauthorized users. As there are many deviations and steps involved, security is also not neglected. In future, the security of real-time data stream among the distributed network will be the area of research. Signifying the efficacy of the proposed DNA encryption method with and outsourced evaluation system for a large number of transactions will be done in future research in terms of privacy.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Subashanthini M, Pounambal M. Three Stage Hybrid Encryption of Cloud Data with Penta-Lyer Security for online business users. Inf. Syst E-Bus Manag. 2020 Sep; 18(3): 379-404.
2. Pavithran P, Mathew S, Namasudra, Lorenz P. A novel cryptosystem based on DNA cryptography and randomly generated mealy machine. Comput Secur. 2021 May;104(1).
3. Xie S, Mohammady M, Wang H, Wang L., Vaidya J, Yuan H. A Generalized Framework for Preserving Both Privacy and Utility in Data Outsourcing. IEEE Trans. Knowl. Data Engg. 2021 May.
4. Leelavathy L, Sugumar R. EUDSFDNA: Enhanced User Data Security Framework using DNA Cryptography in Cloud Computing Environment. J. Shanghai Jiaotong Univ. (Sci). 2020 Sep; 16(9): 568-74.
5. Kannadasan R, Basha MS, Emerson IA. Survey on molecular cryptographic network DNA (MCND) using BIG DATA. 2nd International Symposium on Big Data and Cloud Computing (ISBCC). Procedia Comput. Sci; 2015;50:3-9.
6. Reza M, Torkaman N, Kazazi NS. A method to encrypt information with DNA-based cryptography. International IJCSDF. 2015 Jan; 4(3): 417-26.

7. Majumdar A, Biswas A, Baishnab KL, Sood SK. DNA Based Cloud Storage Security Framework Using Fuzzy Decision-Making Technique. KSII Trans. Internet Inf. Syst. 2019 July; 13(7):3794-819.
8. Xiao W, Hu J. Paradigm and Performance Analysis of Distributed Frequent Itemset Mining Algorithms Based on MapReduce. Microprocess Microsyst. 2020 Jan; 82(5).
9. Kahate A. Cryptography and Network Security. 3rd ed. New Delhi. 2016 July.
10. Hatem MB, Dieaa I. Nassr. DNA-Based AES with Silent Mutations. Arab J Sci Eng. 2018 Sep; 44: 3389-3403.
11. Bhoomika R, Amish Desai. Privacy Preserving Heuristic Approach for Association Rule Mining in Distributed Database. IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems. ICIIECS'15; 2015 Mar; Coimbatore.
12. Lichnu Li, Rongxing Lu, Kim-kwang RC, Datta A, Shao J. Privacy Preserving Outsourced Association Rule Mining on Vertically Partitioned Database. IEEE Trans. Inf. Forensics Secur. 2016 Aug; 11(8): 1-15.
13. Kundalwal MK, Chatterjee K, Singh A. An Improved Privacy-Preservation Technique in Health-Cloud. ICT Express. 2019 Sep; 5(3):1-7
14. Leung CK, Elias JD, Minuj SM, Jesus AR, Cuzzocrea A. An Innovation Fuzzy Logic-Based Machine Learning Algorithm for Supporting Predictive Analytics on Big Transportation Data. IEEE Int. Conf. Fuzzy Syst. 2020 July; Glasgow, UK: IEEE. 19-24
15. Anup R, Dhananjay AJ. Market Basket Analysis: Case Study of a Supermarket. Adv. Mech. Eng. 2020 June.727-34.
16. Attwood TK, Blackford S, Brazas MD, Davies A, Schneider MV. A Global Perspective on Evolving Bioinformatics and Data Science Training Needs. Brief. Bioinformatics. 2019 March; 20(3):398-404.
17. Asadianfam S, Kolivand H, Asadianfam S. A New Approach for Web Usage Mining Usage Mining Using Case based Reasoning. SN Appli. Sci. June 2020;2(7):1-11.
18. Domadiya N, Rao UP. Privacy Preserving Distributed Association Rule Mining Approach on Vertically Partitioned Health Care Data. Elsevier. Procedia Comput. Sci. 2019 Jan; 148: 303-12.
19. Kalia P, Bansal D, Sofat S. Privacy Preservation in Cloud Computing Using Randomized Encoding, Wirel. Pers. Commun. 2021 June; 120:2847-59.
20. Patel T, Patel V. Data Privacy in Construction Industry by Privacy-Preserving Data Mining (PPDMP) Approach. Asian J. Civ. Eng. 2020 Feb; 21:505-15.
21. Chon KW, Kim MS. BIGMiner: A Fast and Scalable Distributed Frequent Pattern Miner for Big Data. Cluster Comput. 2018 Sep; 21(1):1507-20.
22. Tai CH, Philip SY, Ming-Syan Chen. K-Support Anonymity Based on Pseudo Taxonomy for Outsourcing of Frequent Itemset Mining. KDD'10 Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; July 25-28; Washington: United States: Association for Computing Machinery. 2010 Jul. 473-82.
23. Kukade P, Tale R, Thakre S. A Two-Way Encryption for Privacy Preservation of Outsourced Transaction Database for Association Rule Mining. Int. J. Sci. Res. 0218 Mar-Apr; 4(5): 276-85.
24. Domadiya N, Rao UP. Privacy Preserving Association Rule Mining for Horizontally Partitioned Health Care Data. Indian Acad. Sci. 2018 Aug; 43(8):1-9.
25. Suba S, Christopher T. An Efficient Frequent Pattern Mining Algorithm to Find the Existence of K-Selective Interesting Patterns in Large Dataset Using SIFPMM Int. J. Appl. Eng. Res. 2016 May; 11: 5038-45.
26. Patel T, Patel V. Data Privacy in Construction Industry by Privacy-Preserving Data Mining (PPDM) Approach Asian. J. Civ. Eng. 2020 Feb; 9(5): 339-47.
27. Nomura K, Shiraishi Y, Mohri M, Morii M. Secure Association Rule Mining on Vertically Partitioned Data Using Private-Set Intersection. IEEE Access. 2020 Aug; 8: 144458-67.
28. Hua J, Shi G, Zhu H, Wang F, Liu X, Li H. CAMPS: Efficient and Privacy Preserving Medical Primary Diagnosis Over Outsourced Cloud. Inf. Sci. 2020 July; 527: 560-75.
29. Majumdar A, Biswas A, Majumder A, Baishnab KL. A Novel DNA-Inspired Encryption Strategy for Concealing Cloud Storage. Front. Comput. Sci. 2019 Nov; 15(3).
30. Pavithran P, Mathew S, Namasudra S, Lorenz P. A Novel Cryptosystem based on DNA Cryptography and Randomly Generated Mealy Machine. Comput Secur. 2021 May; 104(1).
31. Manjot K, Kiranbir K. Enhanced Security Mechanism in Cloud Based on DNA Excess 3 Codes. Inventive Communication and Computational Technologies. 2020 Jan; 1-12.
32. Min Z, Yang G, Wang J, Kim G. A Privacy Preserving BGN-Type Parallel Homomorphic Encryption Algorithm Based on LWE. J. Internet Technol. 2019 Dec; 20(7): 2189-2020.

33. Stergiou C, Psannis KE. Algorithms for Big Data in Advanced Communication Systems and Cloud Computing. 19th IEEE Conference on Business Informatics 2017 (CBI2017); 2017 July 24-27; Thessaloniki, Greece. Doctoral Consortium: IEEE. 1-16.
34. Stergiou C, Psannis KE, Kim B, Gupta B. Secure Integration of IOT and Cloud Computing. Future Gener Comput Syst. 2018 Jan; 78(3): 964-975.
35. Stergiou C, Psannis KE, Gupta B, Ishibashi Y. Security, Privacy & Efficiency of Sustainable Cloud Computing for Big Data & Iot. Sustain Comput-Infor. 2018 Sep;19:174-84.
36. Stergiou CL, Plageras AP, Psannis KE, Gupta BB. Secure Machine Learning Scenario from Big Data in Cloud Computing via Internet of Things Network. Handbook of Computer Networks and Cyber Security. 2020 Jan;525-54.
37. Stergiou CL, Psannis KE. InFeMo: Flexible Big Data Management Through a Federated Cloud System. ACM Trans. Internet Technol. 2020 Nov; 22(2): 1-22.
38. Denham B, Pears R, Asif Naeem M. Enhancing Random Projection with Independent and Cumulative Additive Noise for Privacy-Preserving Data Stream Mining. Expert Syst. Appl. 2020 Aug; 152(1).113380
39. Xu G, Li H, Ren H, Lin X, Shen XS. DNA similarity Search with Access Control Over Encrypted Cloud Data. IEEE Trans. Cloud Comput. 2020 Jan.
40. Kachouh K, Hariss K, Sliman L, Samhat AE, Alsuliman. Privacy Preservation of Genome Data Analysis Using Homomorphic Encryption. Serv. Oriented Comput. Appl. 2021 Sep; 15(1):273-87.
41. Hagestedt I., Zhang Y, Humbert M. MBeacon: Privacy-Preserving Bacons for DNA Methylation Data. Network and Distributed Systems Security Symposium- NDSS; 2019 Feb 24-27; San Diego, USA; 2019. 1-15.
42. Jain P, Gyanchandani M, Khare N. Big Data Security and Privacy: New Proposed Model of Big Data with Secured MR Layer. Advanced Computing and Systems for Security. 2019 Jan; 883: 31-53.
43. Yin S, Liu J, Teng L. Improved Elliptic Curve Cryptography with Homomorphic Encryption for Medical Image Encryption. Int. J. Netw. Secur. 2019 July; 1-6