Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 3, 2604-2623
https://doi.org/10.28919/jmcs/5573
ISSN: 1927-5307

CODON MODULO CRYPTOSYSTEM FOR PRIVACY PRESERVATION OF VERTICALLY PARTITIONED OUTSOURCED DATA

M. YOGASINI[1,†,*], B.N. PRATHIBHA[2]
Department of Computer Science and Engineering,
Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India
[2]Department of Computer Science, G. Venkataswamy Naidu College, Thoothukudi, Tamilnadu, India

Abstract: Data passes crossways the cloud by the methods of assorted way. It is fundamental to safeguard the data from unapproved users to access the information in any structure. The information refuge is guaranteed by changing a plain text into an incomprehensible configuration by encoding text utilizing cryptographic calculations and these techniques are espoused for scrambling the text to made sure about their data from aggressors to guarantee data protection. Deoxyribo Nucleic Acid (DNA) is an encryption method utilized to provide security to the distributed computing data. In this paper Codon Modulo Cryptography-based Algorithm for vertically partitioned data is applied for privacy concerns. Affiliation Rule Mining and Frequent Itemset strategies are applied to aggregate the Association Rules among the Frequent Items in a scrambled exchange of vertically partitioned information base. The exhibition of Rule Mining calculations such as Apriori, FP-Growth and Eclat with the proposed Codon Modulo algorithm is contrasted with the conventional Homomorphic Encryption.

Keywords: association rule mining; frequent itemset mining; Eclat; Apriori; FP-growth; encryption; codon cryptography.

2010 AMS Subject Classification: 68P25.

[*] Corresponding author
E-mail address: yogaaarudhra@gmail.com
[†] Research Scholar, Register Number: 18124012282038

Received February 19, 2021

1. INTRODUCTION

Data is aggregated as a combined database in the cloud by a few information proprietors. Data insurance is taken into account as the most basic issue in the communication framework because of the flourishing development of transmission applications. While moving this information, this data is gone through unreliable channels. The issue happens while moving this data through uncertain channels to be specific hacking. Cryptography is one of the huge strategies to guarantee secrecy, information trustworthiness in the data conveyed. These strategies are utilized to offer safety to significant data with the end goal that solitary approved individuals can interpret the data. Cryptography [1] is defined as an art of achieving safety by encoding information. Here, method expressions like Encryption and Decryption came into survival. Encryption is characterized as the strategy for encoding plain data to encode text and decryption is characterized as the converse technique of encryption.

Data examination strategies, to be specific Association Rule Mining and Frequent Itemset Mining are used for distinguishing consistently co-happening data. Uses of data examination methods are specifically medical services, prediction, market basket investigation, bioinformatics and web utilized mining. For shielding the value-base data, connections and concealed examples are used. Everything in the value-based information base has a Unique Transaction Id termed as UTID. Let Z be a frequent itemset in an exchange only when Support (Z) ≥ ST.

Support is characterized as the number of things that arise regularly in the exchange. ST is characterized as the Support Threshold demarcated by the information proprietor. $(A \rightarrow B)$ is an affiliation rule where A and B are two distinct item sets which illustrates that, each time X take place in a transaction Y also occur in the same transaction. Confidence of $(A \rightarrow B)$ is demarcated as the possibility of how possible B is obtained whenever A is attained. Frequent items in an exchange are achieved relying upon the threshold standard which authenticates to apprehends association rules on every frequent item. At the point when a bunch of frequent items has been enforced in an exchange, affiliation rules are made.

A novel DNA algorithm to protect the cloud data is implemented in this work. During exchanges, a substitution cipher is adopted for scrambled things, to disguise the information proprietor’s data in opposite to frequency analysis bouts. Affiliation Rule Mining and Frequent Itemset Mining are enforced to accumulate association rules among frequent items in a scrambled exchange of vertically partitioned database. Apriori, Eclat and FP-Growth affiliation rule mining calculations are utilized for refining the association rule with various k-values, where k signifies the things in exchange. In this paper, a protection safeguarding procedure is intended for vertically partitioned information by smearing DNA cryptography-based algorithm.

2. Related Works

There is plenty of works completed associated with the privacy preservation of both horizontal and vertical divided cloud information. The literature provides an overall though of the methods that could be pertained for security safeguarding with their focal points and burdens. A portion of refereed works here are in short.

P. Kukade et al [2] adopted Paillier Homomorphic Encryption which provides good security than prevailing encryption algorithm for Transactional Information base that is proper for outsource affiliation control mining, FP-Growth assessment is employed to find affiliation rule mining over Apriori which has a supreme performance. The proposed background has more execution with respect to time and leads to less security.

V. Redekar et al [3] proposed an enhanced Rob Frugal encoded algorithm which supported on accumulating weighted support in unique things support exchanges to decrease the phony exchange table information and matrix is produced to diminish the capacity overhead. They also proposed Elliptic Curve Diffie Hellman (ECDH) key exchange procedure after the Rob Frugal algorithm to beat the speculating assault. The projected system lessens the number of phony patterns and upgrades the safety level by adding the weighted support to novel support of objects for outsourced information with less complication.

A sufficient and convincing scattered estimation namely Fast Distributed Mining (FDM) was proposed by K. Mariappan et al [4] to mine association rules. Apriori computation was adopted to discover every single massive control among items in a information base of a transaction. This algorithm is to determine affiliation among several measures of data. The data can be kept protected and the information emission is in a smaller amount. This algorithm is not applicable for vertically partitioned information base.

A.K. Jumaah and S. Janabi et al [5] proposed a novel procedure to conquer the disadvantages of the two existing algorithms specifically Increase Support of Left (ISR) and Decrease Support of Right (DSR). This algorithm is adopted for concealing subtle rules grounded on ISL and DSR. The investigational result has indicated that the projected algorithm successfully lessens the side effects, decline the certainty of the subtle rules and generates various novel principles between things and manages the Left-Hand Side (LHS) and Right-Hand Side (RHS) together as per the proportion among them and it chooses the exchange with the least weight for altering unique data set. This approach is not applicable for huge dataset.

K. Agrawal and V. Tewari [6] proposed a collaborative Privacy-Preserving Data Mining (CPPDM) approach for outsourced information, which guarantees that the information is put away handled, and shared without disregarding the client protection with the help of anonymization and encryption methods. Different information proprietors can securely outsource their information. MapReduce System is executed to improve security.

X. Wang, L. Bai et al [7] adopted fully homomorphic encryption for security assurance of eHealth Record in the cloud. The system can ensure the privileges of patients and specialists as far as possible, and simultaneously forestall discretionary conduct of one or the other side, Electronic wellbeing security fortification and information preparing are dependent on complete homomorphism idea to keep monstrous clinical information in eHealth framework which encompasses various subtle data electronic wellbeing records of clients.

H.M. Bahig and D.I. Nassr [8] presented a novel version of Advanced Encryption Standard (AES), known as DNAES, grounded on deoxyribonucleic corrosive (DNA) by stowing binary information systems with silent mutations. The novel method utilized to plan the DNAES-cipher can be summed up to several chunk ciphers. This method depends on DNA grouping, where one gram of DNA accumulates around 106 TB of information. IN DNA, molecules are used to implement Hamiltonian path problem, parallel operations, and computations. It has a similar security level as AES but not more than AES.

C. Gritti, W. Susilo et al [9] proposed a novel system termed Ciphertext Policy DNA-Based Encryption (CP-DBE) to encipher information utilizing the DNA arrangements of the despatcher to decipher information utilizing the DNA arrangements of the beneficiary. Set of intersection procedure and generation of token delivered at the fixed time are two additional highlights to unscramble the code text. The CP-DBE protects in opposition to malicious adversaries and collusion resistance in the arbitrary oracle prototype, beneath the Decisional Parallel Bilinear Diffie-Hellman Exponent supposition.

An innovative DNA-based Privacy-Preserving (DNAPP) system was introduced by W.M. Abed [10] that guarantees strong confirmation, concealment, message trustworthiness, and guaranteeing high users' security. Acceptable security highlights of this procedure are high complexity of o(n!), light-weight, versatile, least overhead. Since the conveying parties can decide the key domestically and autonomously, there is no need for cryptography key interchange among them. This system requires additional examinations and assess its presentation in contrast to various safety assaults.

A new security assurance procedure was adopted utilizing DNA encryption and a hyperchaotic system by S. Cheng, L. Wang et al [11]. The author projected an incorporated deep hashing calculation to extricate highlights dependent on the coordinated deep network prototype and they study hash codes dependent on the projected highlights. In the record encoding measure, the KNN method is utilized to scramble the record to guarantee refuge with efficiency.

S.K. Sood [12] proposed an edge work containing three cryptographic constraints such as Confidentiality (C), Availability (A) and Integrity (I) to ensure the information with different estimates, for example, the SSL (Secure Socket Layer) and MAC (Message Authentication Code). Proposed technique accomplishes the accessibility, consistency and trustworthiness of information navigating through proprietor to cloud and cloud to client. In addition to that it likewise gives greater adaptability and ability to fulfil the new need of the present intricate and different organization and furthermore empower the client to recover records from cloud via looking over an encoded information.

M. Najaftorkaman and N. S.Kazzai [13] proposed a novel strategy to scramble information by utilizing DNA-based cryptography based on quantum and DNA cryptography. The exemplary Vigenere figure DNA cryptosystem ideas were discussed and this technique was safer because it has two layers of security, which are computational and genetic security. Since this strategy was a novel technique, they need some improvement. Furthermore, amenities like DNA chips and robots are required for exploratory arrangements.

Field Programmable Gate Arrays (FPGA) was applied by S. Sasikumar and P. K. Kumar [14] for information refuge. In this paper, the idea of Quantum Cryptography (QC) and DNA-based calculation was utilized. DNA based calculation was adopted to create a key for scrambling and unscrambling message. When contrasted with the current encryption framework, the proposed framework is computationally more productive. This is, despite the fact that QC and DNA cryptography are in their beginning phase, they give the best security and are quicker to execute.

Haoyuan Li et al [15] Proposed a Private FP-growth (PFP-growth) calculation, that consists of a pre-processing stage and a mining phase. Dynamic decrease technique to animatedly lessen the measure of noise added to ensure protection during the mining progression. Formal security investigation and the consequences of wide analyses on genuine datasets illustrate that the PFP-growth system is time-productive and can accomplish together virtuous efficacy and great protection.

F. Giannotti, V.S. Lakshmanan et al [16] recommended an assault model dependent on foundation information on protection conservation and devise a plan for security safeguarding outsourced mining. They built up an encryption system that empowers formal security assurances to be demonstrated, and to approve this model over huge scope genuine exchange information base. The authors proposed encryption conspire, called Rob-Frugal, that depends on 1-1 substitution ciphers for things and adding counterfeit exchanges to ensure the information. In future, Rob-Frugal calculation is reached out to limit the number of fake patterns.

H. Bae, S. Min et al [18] proposed DNA Steganalysis model to distinguish concealed messages paying little heed to the concealing position. This model customs unsubstantiated pre-training of a sequence-to-sequence autoencoder to acquire the intrinsic protocols of DNA arrangements and their inner design of unmodified genome groupings. They also introduce RNN-CNN-based steganalysis model which does not need earlier information on the utilized steganography calculation. The proposed model can identity concealed message autonomous of DNA steganography calculations. To decode the shrouded messages to the current finding abilities.

Kannadasan et al [19] have utilized a DNA-based encryption procedure in enormous information. At the point when an enormous amount of data is to be saved by applying huge information then encryption strategy is applied. A DNA encoding table with PHP language is utilized for the encryption cycle.

3. METHODOLOGY

The proposed technique presents security safeguarding for outsourced vertically segment information base with the proposed Codon Modulo cryptography calculation. The framework model of the proposed methodology in the cloud database is portrayed in Fig 1. In the pre-processing phase, the ciphertext substitution technique has been utilized for the protection of data in the cloud. The actual fictitious data of both the information proprietors are encoded with the new cryptosystem method. The support threshold values are predetermined by the dispatcher for assessing frequent items in exchange.

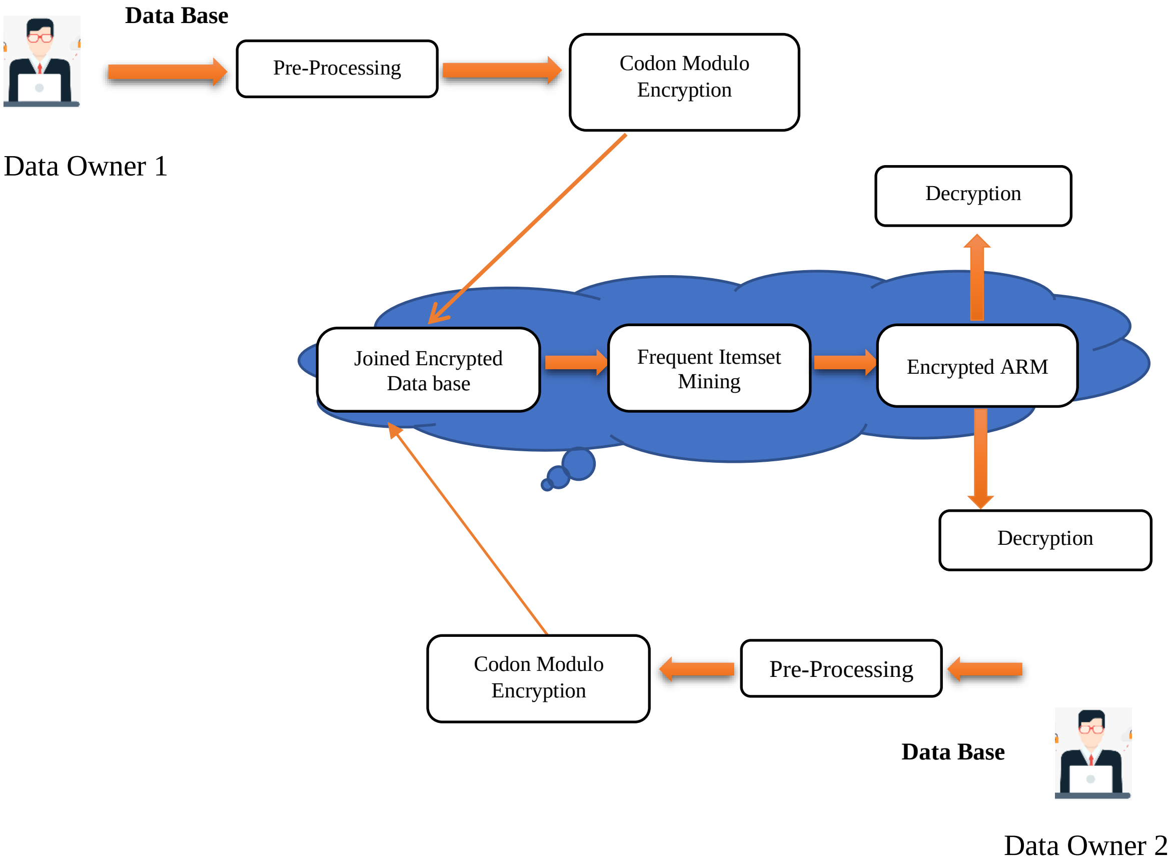

FIGURE 1. System Model for Proposed Method in Cloud

3.1. Pre-processing

3.1.1. Substitution Cipher

In a Substitution cipher, any character of plain content from the specified static set of characters is subbed by some other character from a similar set contingent upon key. Information proprietor's message is scrambled utilizing the substitution cipher that is foremost to the outsourcing technique. Each item in the exchange has an equal replacement text. A substitution cipher is an issue to frequency testing assault.

Table 1: Cipher text substitution

Plaintext	Biscuit	Milk	Bread
Ciphertext	1100	0010	1101

A substitution cipher is reliant upon recurrence investigation assault if the frequencies of message units are unique. Frequency analysis has been utilized to breakdown traditional codes like substitution ciphers. Aggressors with some information on the occurrence of message units or unit gatherings can recuperate some plaintext through recurrence investigation. To secure recurrence investigation assault, fake items can be added to conceal item recurrence.

3.1.2. Hash Function in Cryptography

Cryptographic hash work hf ( ) is a distinctive sort of hash work utilized in cryptography and this function is applied which maps self-assertive strings of information length of fixed yield in deterministic arbitrary technique. The hash function has the characteristics of pre-picture opposition and crash obstruction. It is a single direction work. SHA-1 [22] and SHA-2[23 base paper ref] are frequently utilized cryptographic hash functions.

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

3.1.3. Fictitious Transaction

Information proprietors add counterfeit things to their non-public message to safeguard their data. Counterfeit things are autonomously added into the encoded exchanges in order to enhance the fortification. Information proprietor denoted their exchange whether it is genuine exchange or phony with a flag variable.

Y fictitious exchanges are implanted among every two unique exchanges, where Y is a random variable. Encrypted Realness Values (ERV) are labelled by the information proprietor in their transaction and all ERV values are directed to the cloud. The flag values is set as Zero for counterfeit exchange and one for authentic exchange.

3.2. Codon table

DNA is an exceptionally extensive atom finished by a long chain of nucleotides. Nucleotides are found generally in the cell core. Every nucleotide contains deoxyribose, which has a place with a phosphate gathering and a nitrogenous base, to be specific Adenine A, Cytosine C, Guanine G and Thymine T. These sorts of nitrogenous bases are arranged in all nucleotide, which separates one nucleotide from other. In addition, a long chain arrangement of nucleotides is printed as a succession of nitrogenous bases with the connected appearance in the nucleotides. The arrangement of nitrogenous bases frames the hereditary code of cells. The hereditary code contains three sequential nucleotides, which is known as a codon for instance TTT, CAG, ACT. Aside from TAA, TGA and TAG, every codon of DNA utilizes on the production of one of 20 amino acids. TAA, TGA and TAG show a finish of an arrangement of codon. An association of coupled amino acids and their request for game plan are called as protein. Therefore, a progression of codon in DNA particles are coded in a particular sort of protein and figure a quality. One quality can contain up to at least 1,000 codons.

Table 2: Amino acids with their equivalent codons

Amino acid	Codons
Isoleucine	ATA, ATC, ATT
Leucine	CTA, CTC, CTG, CTT, TTA, TTG
Valine	GTA, GTC, GTG, GTT
Phenylalanine	TTC, TTT
Methionine	ATG
Alanine	GCA, GCC, GCG, GCT
Glycine	GGA, GGC, GGG, GGT

The number of attainable codons is equal to $3^3 = 27$. The number of codons pragmatic to generate amino acids is equal to 24, which is computed by subtracting three codons TAA, TGA and TAG from 27. The number of attainable amino acids is 20, which is lesser than 24 dissimilar codons to produce the same amino acid. Table 2 epitomizes all 24 codons with their equivalent amino acids.

3.3. The proposed Codon Modulo Cryptosystem

Let us consider DNA sequence as input to the encoding technique. The input sequence is transformed into binary bits. Then the binary input sequence is divided into N number of sets, where each set is a composition of four bits. Here, encoding is achieved on each set distinctly using ACGT codons table as shown in Table 3.

Table 3: Codons table for ACGT

A_Codans	ATA, ATC, ATT ATG ACA, ACC, ACG, ACT AGC, AGT AAC, AAT AAA, AAG AGA, AGG
C_Codans	CTA, CTC, CTG, CTT, TTA, TTG CCA, CCC, CCG, CCT CAA, CAG CAC, CAT CGA, CGC, CGG, CGT
G_Codans	GTA, GTC, GTG, GTT GCA, GCC, GCG, GCT GGA, GGC, GGG, GGT GAA, GAG GAC, GAT
T_Codans	TTC, TTT TGC, TGT TCA, TCC, TCG, TCT TAC, TAT TGG

Table 4: Randomly Shuffled Codon Table for ACGT

A_Codans	AGT, ACC, AAG AGC AAC, ATC, AGA, ATG AAA, ACC AGC, ATT ACT, ATA ACG, AAT
C_Codans	CAA, TTA, CCG, CGT, CCC, CGG CTA, CAG, CAT, CTG CTT, CGC CGA, CAC CCT, CCA, CTC, TTG
G_Codans	GCT, GGT, GCC, GAA GAG, GTC, GGC, GAC GAT, GCA, GTA, GTT GTG, GCG GGG, GGA
T_Codans	TCC, TGG TAT, TCT TAC, TGC, TTT, TCA TCG, TTC TGT

The original codon in the codon table shown in Table 3 is randomly shuffled for encryption purpose. The shuffled codon table is shown in Table 4.

3.3.1. Encryption Process

Every item in exchange is supplanted with the codons in the codon table. For encoding measure, DNA succession is going about as information. Division and modulo capacities are to be accomplished with the complete number of codons for the chose thing in exchange. After obtaining the quotient and remainder from the respective functions, check the codon in the codon table which matches with the corresponding results. The coordinated codons are chosen for encryption.

Let Q and R be the Quotient and reminder respectively and I be the selected item.

$$Q = I \text{ div number of codons} \qquad (1)$$

$$R = I \text{ mod number of codons} \qquad (2)$$

Let us consider an example. Let 169th item in a transaction is taken for encryption. Division and modulo function are to performed on the selected item with the total number of codons. The Q and R value will be obtained by using Equ (1) and Equ (2). The value of Q will be 2 and the value of R will be 47. The corresponding codon value in the codon table is taken for encryption. The chosen codon form the shuffled codon table dependent on the aftereffect of Equ(1) and Equ (2) is ACCGTG. The primary letter of the First codon and the main letter of the subsequent codon is joined and the second letter of the First codon and the second letter of the subsequent codon is consolidated and so on to frame the scrambled information. The supplanted codon of ACC for 2 and GTG for 47. The combined codon combination is AGCTCG. The codon equal for the item is AGCTCG.

3.3.2 Decryption Process

The two codons are separated as concatenating First, Third, and Fifth positions of the combined codon AGCTCG and concatenating the second, fourth and sixth positions of the combined codon. The separated codons are ACC and GTG respectively. The ACC's codon position in the shuffled codon table will be found. Since ACC is in the second position in the table, Consider the value of Q as 2. Similarly, the GTG's codon position in the table is 47. Hence the value of R is 47.

This process will be performed by multiplying Q with the total number of codons and add the Remainder with the obtained result to get the decrypted data. This function will be denoted as

$$(Q \text{ X number of codons}) + R = \text{original data.} \quad (3)$$

With the above example, The Q value 2 is multiplied with the number of codons in the shuffled codon table then the result will be added with the R value 47 which leads to the decrypted data. The result of the Equ (3) is the selected item in a transaction. The resultant decrypted text equivalent to the codon AGTACA is 169.

4. OUTSOURCED ITEMSET MINING

Frequent Itemset Mining (FIM) is the procedure utilized generally in the field of information mining like human services framework. A frequent itemset is an itemset aim at finding uniformities in the shopping behaviours of customers in supermarkets and on-line shops, etc. Frequent itemset mining is an interesting division of data mining. An aggregation of one or additional points is coined as Itemset.

In FIM for the most part the intriguing affiliations and connections between itemset in transactional and relational databases are found and it shows the things show up together in a transaction or relation. FIM is used to help cross-selling (proposal of other products) and product heap. The purpose of FIM is to detect all subset of items that occur together in many transactions and discover the association rules that relate the occurrence of one set of items with that of another set of items in the transaction database.

Let I be Items, X be Itemset, T and D are the set of transactions and Database respectively, then $X \subset T$. k is an itemset of length k, where length is the number of elements in the itemset. The support of an itemset X is defined as $X = \{T \in D \mid X \subset \mathrm{T}\}$. An itemset X is called frequent when $support(X) \geq minsup$.

5. ENCRYPTED ASSOCIATION RULE MINING

Association Rule Mining (ARM) is used to discover patterns in a transaction based on the result of frequent itemset mining. The goal of rule mining is to forecast the incidence of a specific item in a transaction and it adopts machine learning models for analyzing the co-occurrence of data in the database. Transaction secrecy is preserved by employing association rules. When association rule mining is performed on the encrypted information, the privacy of items is preserved.

Transaction secrecy preserves by employing association rules. When association rule mining is performed on the encrypted information, the privacy of items is preserved. Association Rule Candidate $X \to Y$ that satisfies $X \cap Y = \emptyset$ and $X \cup Y$, where X, Y and $X \cup Y$ are seemingly frequent item-sets.

Here, in vertically partitioned database the data owners possess one or more attributes in the joint database [17]. The description of dilemma of mining association rules can be defined as T=$\{l1, l2, \ldots, ln\}$, which is a set of literals named as items. A collection of transaction is represented by L, where for each transaction L is a set of items such that L⊆T. A unique identifier named as Transaction ID (TID) is integrated with each transaction. Moreover, transaction L contains P, which is a set of some items in L, if$P \subseteq$L. An association rule is denoted as $P \Longrightarrow Q$, where$P \subseteq$ T,⊆ T, and$P \cap Q = \phi$. The rule$P \Longrightarrow Q$subsists in the transaction set L with confidence value c. So, c% of the transactions in L that contain P also contains Q. The rule$P \Longrightarrow Q$has support value s in the transaction set L. When s% of the transactions occurs, then T contains $P \cup Q$. If the support value of an itemset P is greater than or equal to the minimum support threshold, then it is named as frequent itemset.

Finding frequent itemset in a transaction and generating strong association rules among the frequent items are the two steps in involved in ARM. This is performed on the basis of support and confidence value in a transaction.

Support is the frequency of items and the combination of items bought in a transaction. The items with less support value (less frequency value) are filtered.

$$Support = freq\frac{(A,B)}{N}$$

Where N represents the total number of transactions present in a transactional database.

Confidence indicated how often items occurred together. Confidence states the relationships of the items have been found to be true. Confidence of the rule $(A \to B)$ is the transaction which contain A also contain B. Formula to calculate confidence is Confidence $(A \to B)$= AUB/A

Confidence is defined as the relationships of the items have been established to be correct. Confidence of the rule $(A \to B)$ is the transaction which include A also include B. It is represented as shown in Eq.(3).

$$\text{Confidence } (A \to B)\text{= } AUB/A$$

6. EXPERIMENTAL EVALUATION

The experimental evaluation is performed on the retail and the pumsb data set, which is an open-source data set. This retail data set contains 88,162 transactions. It also has retail market basket information with a Belgian retail store. This pumsb data set contains 49,046 transactions. This also contains census data for population and housing. Complexity calculation is achieved utilizing these datasets with the proposed calculation. This work likewise emphasis around looking at the cloud execution period of three mining calculations, to be specific Apriori-DNA, Eclat-DNA and FP-Growth DNA on vertically partitioned exchanges. The execution cloud time of four unique exchanges with various k-values is assessed where k is the number of items in an exchange. The concert of these three algorithms is assessed autonomously for the predefined support threshold value of items in the exchange set with the assistance of Python. The transaction set taken for the examination was 5000, 10000, 15000, 20000 with dissimilar k-values.

Table 5: Cloud Execution Time with different k-Values for Pumsb Data-Set using 5000 and 10000 Transactions

Transaction	**5000**						**10000**					
K values	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA
k=10	3.967	2.994	4.108	4.968	6.025	5.547	6.017	5.101	7.257	5.297	8.127	6.128
k=20	4.957	3.831	5.158	5.968	7.958	6.011	6.954	5.984	8.124	6.285	9.287	7.107
k=30	5.928	4.988	6.981	6.451	8.527	7.847	7.147	6.247	9.058	7.368	10.057	8.047
k=40	6.827	4.154	7.024	7.651	9.287	8.001	8.107	7.348	9.987	8.197	11.198	9.198

Table 6: Cloud Execution Time with different k-Values for Pumsb Data-Set using15000 and 20000 Transactions

Transaction	**15000**						**20000**					
K values	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA
k=10	7.957	5.978	8.857	7.017	9.125	8.718	8.998	7.017	9.139	8.258	10.928	9.054
k=20	8.017	6.287	9.387	8.157	10.028	9.117	9.017	7.927	10.528	9.983	11.051	10.085
k=30	9.274	7.374	10.128	9.574	11.231	10.218	10.257	8.120	11.112	10.127	12.027	11.147
k=40	10.198	8.987	11.598	10.167	12.543	11.312	11.827	9.968	12.008	11.085	13.991	12.017

From Table 5 and Table 6, it is seen that the running time changes with expanding estimations of k. In Table 5, while considering about k value as 30 with 10000 exchanges, the outcome of Eclat with DNA is 6.247, which has least cloud execution time. In Table 6, while thinking about k values as 20 with 20000 exchanges, the outcome acquired of Eclat with DNA encryption is 7.927, which has least cloud execution time.

Table 7: Cloud Execution Time with different k-Values for Retail Data-Set using 5000 and 10000 Transactions

Transaction	**5000**						**10000**					
K values	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA
k=10	3.294	2.147	4.847	3.387	5.121	4.746	4.487	3.015	5.978	4.358	6.157	5.368
k=20	4.108	3.014	5.158	4.154	6.964	5.027	5.987	4.317	6.482	5.028	7.587	6.425
k=30	5.148	4.652	5.948	5.745	7.514	6.247	6.148	5.498	7.475	6.428	8.547	7.385
k=40	6.014	5.278	6.887	6.017	8.124	6.987	7.471	6.357	8.147	7.289	9.578	8.118

Table 8: Cloud Execution Time with different k-Values for Retail Data-Set using15000 and 20000 Transactions

Transaction	**15000**						**20000**					
K values	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA	Eclat Homo	Eclat DNA	Apriori Homo	Apriori DNA	Fpgrowth Homo	Fpgrowth DNA
k=10	5.458	4.512	6.147	5.074	7.478	6.149	6.147	5.276	7.148	6.247	8.074	7.168
k=20	6.147	5.142	7.294	6.247	8.941	7.358	7.998	6.048	8.059	7.014	9.147	8.248
k=30	7.457	6.024	8.247	7.147	9.248	8.011	8.105	6.982	9.487	8.187	10.784	9.144
k=40	8.947	7.247	9.149	8.034	10.687	9.345	9.854	8.142	10.183	9.017	11.475	10.387

From Table 7 and Table 8, it is observed that the running time changes with increasing values of k. In Table 7, while considering k value as 40 with 5000 transactions, the result obtained in Eclat with DNA encryption is 5.278, which has minimum cloud execution time when compared to Apriori and FP-Growth with respect to homomorphic encryption. In Table 8, while considering k value as 30 with 20000 transactions, the result obtained is Eclat with DNA encryption is 6.982, which has minimum cloud execution time when compared to Apriori-Homo and FP-Growth Homo.

The cloud's running time increments with k for the pumsb dataset, however scarcely changes for the retail dataset. The growth in running time for pumsb dataset is because of the rise in fake information. Nonetheless, the retail dataset is thick, and the backings are now high without including fake information. Consequently, adding more fake information scarcely changes the quantity of apparently frequent itemsets and their backings. The cloud's running time scarcely changes for the pumsb dataset because the pumsb dataset is extremely opaque. The increase in running time for pumsb dataset is due to the increase in pretended information.

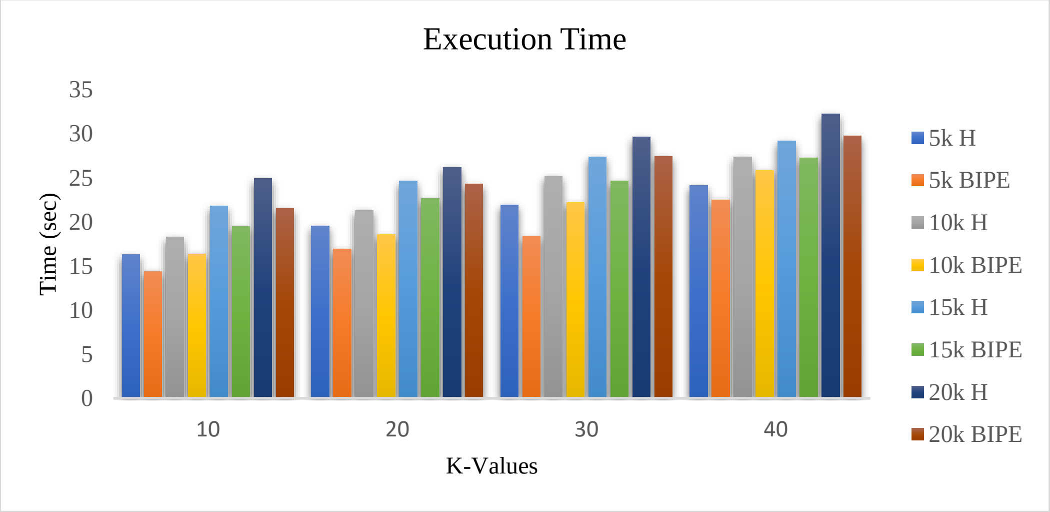

FIGURE 2: Data owner Execution time with different k values for pumsb data set

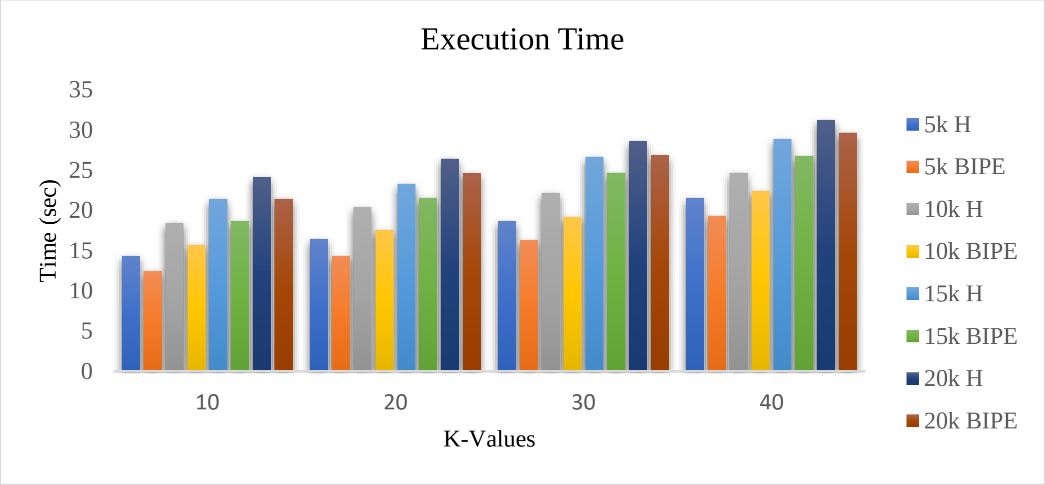

FIGURE 3: Data owner Execution time with different k values for retail data set

The information proprietor’s execution time of various transactions with different k-values for both pumsb and retail dataset in Fig 2 and Fig 3 respectively. Eclat algorithm with the proposed method has less data owner execution time in different size of transactions.

Information proprietor’s running time does not increment with k-values by the same token . In particular, the information proprietor’s running time is overwhelmed when to run this algorithm, which is barely exaggerated by the variations in the values of k.

7. Conclusion

In this paper, a privacy-preserving subcontracted frequent itemset mining scheme for the vertically apportioned information base. This scheme authorizes the information proprietors to subcontract mining job on their combined information in a security protecting way. Because of this scheme, a privacy-preserving subcontracted affiliation rule mining solution was erected for vertically divide information bases. An effective DNA based encryption and a protected outsourced assessment system were introduced in this paper. These solutions additionally guarantee the security of the mining outcomes from the cloud.

The experimental consequences portray that the time taken for encoding illustrates less time contrasted with the existing system. Information proprietors when choose to outsource their information base to the cloud prerequisite a significant level of security without negotiating performance can adopt this novel proposed method. Comparison of three mining algorithms is adopted and give proficient outcomes. The outcomes are portrayed precisely with less computation time. This proposed strategy has diminished the execution time to 21 % for 5000 and 10000 transactions and 17% for 15000 and 20000 transactions while considering amalgamation with the homomorphic encryption calculations in the Retail dataset. Likewise, the execution time has diminished to 27% for 5000 and 10000 transactions and 19 % for 15000 and 20000 transactions for the pumsb dataset.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] A. Kahate, Cryptography and Network Security: 3rd edn. Tata McGraw Hill Education Private Limited, New Delhi, 2013

[2] P. Kukade, R. Tale, S. Thakre. A Two-Way Encryption for Privacy Preservation of Outsourced Transaction Database for Association Rule Mining, Int. J. Sci. Res. Sci. Technol. 4 (2018), 276-285.

[3] V.R. Redekar, K.N. Honwadkar, Privacy-Preserving Mining of Association Rules on Cloud by Improving Rob Frugal Algorithm, Int. J. Res. Inform. Secur. 5 (2015), 32-38.

[4] K. Mariappan, G.V. Sriramakrishna, M.M. Selvam, G. Suseendran, Data Secure in Horizontally Distributed Database Using Apriori Algorithm, Int. J. Eng. Technol. 7 (2018), 146-149.

[5] A.K. Jumaah, S. Al-Janabi, N.A. Ali, An Enhanced Algorithm for Hiding Sensitive Association Rules Based on ISL and DSR Algorithms, Int. J. Comput. Network Technol. 3 (2015), 84-89.

[6] K. Agrawal, V. Tewari, Privacy-Preservation in Collaborative Association Rule Mining for Outsourced Data, Int. J. Distr. Cloud Comput. 5 (2017), 30-34.

[7] X. Wang, L. Bai, Q. Yang, L. Wang, F. Jiang, A dual privacy-preservation scheme for cloud-based eHealth systems, J. Inform. Secur. Appl. 47 (2019), 132–138.

[8] H.M. Bahig, D.I. Nassr, DNA-Based AES with Silent Mutations, Arab. J. Sci. Eng. 44 (2019), 3389–3403.

[9] C. Gritti, W. Susilo, T. Plantard, K.T. Win, Privacy-preserving encryption scheme using DNA parentage test, Theor. Computer Sci. 580 (2015), 1–13.

[10] W.M. Abed, A DNA-Based Privacy-Preserving Scheme in Smart-Grid, Int. J. Cryptogr. Inform. Secur. 9(2019), 1-10.

[11] S.-L. Cheng, L.-J. Wang, G. Huang, A.-Y. Du, A privacy-preserving image retrieval scheme based secure kNN, DNA coding and deep hashing, Multimedia Tools Appl. (2019). https://doi.org/10.1007/s11042-019-07753-4.

[12] S.K. Sood, A combined approach to ensure data security in cloud computing, J. Network Computer Appl. 35 (2012), 1831–1838.

[13] M. Najaftorkaman, N.S. Kazai, A Method to Encrypt Information with DNA-Based Cryptography, Int. J. Cyber-Secur. Digital Forens. 4(3) (2015), 417-426.

[14] S. Sasikumar, P. Karthigaikumar, VLSI implementation of DNA cryptography using quantum key exchange, in: 2014 International Conference on Electronics and Communication Systems (ICECS), IEEE, Coimbatore, 2014: pp. 1–5.

[15] H. Li, Y. Wang, D. Zhang, M. Zhang, E.Y. Chang, Pfp: parallel fp-growth for query recommendation, in: Proceedings of the 2008 ACM Conference on Recommender Systems - RecSys ’08, ACM Press, Lausanne, Switzerland, 2008: p. 107.

[16] F. Giannotti, L.V.S. Lakshmanan, A. Monreale, D. Pedreschi, H. Wang, Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases, IEEE Syst. J. 7 (2013), 385–395.

[17] S. Varma, I. Lijip, Secure Outsourced Association Rule Miming using Homomorphic Encryption, Int. J. Eng. Res. Sci. 3 (2017), 70-76.

[18] H. Bae, S. Min, H. Choi, S. Yoon, DNA Privacy: Analyzing Malicious DNA Sequences using Deep Neural Networks, IEEE/ACM Trans. Comput. Biol. Bioinform. (2019), http://data.snu.ac.kr/wp-content/uploads/2020/08/TCBB2020.pdf

[19] R. Kannadasan, M.S. Saleembasha, I. ArnoldEmerson, Survey on Molecular Cryptographic Network DNA (MCND) Using Big Data, Procedia Computer Sci. 50 (2015), 3–9.