IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 7, July 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

Optimization Based Privacy Preserving Data Publishing Model For Cloud Computing

Sahana Lokesh R[1] , Dr. H R Ranganatha[2]

Assistant Professor, Department of CSE, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India

Professor& Head, Department of ISE, Sapthagiri College of Engineering, Bengaluru, Karnataka, India

ABSTRACT: A very important attribute for the information stored inthe cloud platform is data privacy. It is a biggest challenge to share and publish medicalinformationin cloud infrastructure with sensitive information about an individual.So that security and privacy preservation ofpatientsis essential. In this paper,privacy of the patients is preserved withMondrian based k-anonymizationwith a novel Genetic Chimp Optimization Algorithm (GA-ChmpOA).This optimization algorithm utilizes average equivalence value and generalized loss of Information for the calculation of fitness value. This paper also proposes a retrievable data encryption model for overcoming the challenges in cloud computing. Moreover, Genetic-DNA based encryption technique is utilized after anonymization process to give extra protection to the anonymized database.The performance of the proposed privacy preservation technique is evaluated with respect to information losses,privacy and utility.When compared to other techniques, the proposed approach shows better results and it is efficient to preserve the privacy of medical database.

KEYWORDS: Privacy preservation, Anonymization, utility, privacy, Data security, Encryption.

I. INTRODUCTION

Cloud Computing is the network of networks to access computing resources over the Internet, and it is a new computing archetype that provides various services on demand at a low-cost [1].To store, manage and retrieve the information such as individual data, medical records,and financial transactions the cloud environment has been utilized by millions of peoples all over the world.These data may contain sensitive information about individuals. Personally identifiable information (PII) such as emails, quasi-identifier, address, name, date of birth, contact number, zip code, passwords and personal sensitive data such as health, finance, treatment details, medical history etc. are known as sensitive information [2]. Individual's privacy isat risk when publishing the data with such sensitive information.To protect individual's privacy, it is necessary to anonymous the database before store it in the cloud.

The process of excluding PII from data record before transferring the data to the CSP is known as anonymization [3, 4].k-anonymization is one of the most popular anonymization method that guranteeseach data record is indistinguishable from at least k-1 other records [5]. Suppression and Generalization are the two main operations of the k-anonymization process.Though this technique shows better performance in individual'sprivacy preservation, the performance is ineffective for high dimensional and multi-dimensional data [6]. Mondrian based k-anonymization technique can handle multi-dimensional database.In this approach, database is partitioned into high dimensional numerical data [7].

Preserving both the utility and privacy in the anonymous data simultaneously is a main challenge in PPDP. So that, metaheuritic based optimization technique is utilized with Mondrian based k-anonymization technique to improve theprivacyandutility. The optimization algorithm used in this paper is combination of genetic algorithm (GA) [8] and chimp optmization algorithm (ChmpOA) [9]. ChmpOA is inspired by sexual motivation and individual intelligence of chimps in their group hunting. GA is a search heuristic that is inspired by Charles Darwin's natural evolution theory.GA has proven to be a reliable and powerful optimization technique applied to a wide variety of real-world issues of significant complexity.Though the anonymized data can protect sensitive data if it is published in the cloud in the non-encrypted format, it can be easily learned by the public and cloud users. So that, DNA-Genetic based

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

encryption technique [11] is utilized in theproposed approach to protect the sensitive information in the cloud. This encryption technique is more stable because of the genetic and multi-iteration activities.Adult database [12], Indian Liver Patient Records [13], Early Stage Diabetes Risk Prediction Dataset[14] and COVID-19 patient pre-condition dataset[15] are utilized to evaluate the proposed approach.

The rest of this paper is arranged as follows. Section 2 details the literature review. Section 3 explains the proposed methodology. Section 4details experiment and section 5 experimental results. Section 6 concludes this paper.

II. RELATED WORKS

Cloud computing provides resources such as, collaborating servers, virtual data storage, applications and networks. Due to the high amont of information in cloud storage, the secuity of sensitive information is the critical issue [10]. DoS attack, data breeching, security threats, compromised authentication, malicious insider, data privacy, data integrity and vulnerable systems are some of the security issues faced by the cloud computing [16]. In the cloud computing, the access control models are based on task, action, attribute, encryption and usage. To deal with the privacy issues and cloud data integrity lot of works are proposed by many researchers.

2.1 Data anonymization

For the protection of privacy, the k-anonymity model is introduced by [17] has received more attention from researchers. In order to achieve k- anonimity various algorithms such as generalization and suppression are proposed. In [18] k- anonymization is proposed for the preservation of privacy in data collection. Before the data from the owners reaches the user, the data is anonymized in a collaborative and distributive way. This approach is not secure and possible for attacks like attribute disclosure and membership disclosure.In [19], a similar privacy presevation based data publishning model using k-anonymity approach is proposed. Outsider and insider privacy attacks are addressed in this model. With the help of Oracle vault technique, insider attacks are solved. K-anonymity technique is utilized to neutaralize the outsider attacks. To improve the k-anonimity based privacy model, a centralized method known as Mondrian is proposed in [20]. For partitioning the data this algorithm utilizes multidimensional generalization in top-down approach. In [21] three centrailized techniques are introduced to generalize the hierarchies. Three techniques are focused on partitioning and data utility. The number of nodes at each level, distortion ratio and discernibility are determines the efficiency of this approach. Genetic Grey wolf algorithm (GA-GWO) based k-anonymization is proposed in [22] for the privacy based data publishing model. Cavg, GIloss, utility and privacy are the evaluation metrics in this approach.

2.2 Data encryption

GA based data encryption approach is proposed by authors of[23]. A random key is generated in this approach and encyption is performed with genetic operations such as mutation and crossover. Between the key and the plaintext, the XOR operation is performed to generate the ciphertext. The main goal of this approach is random key generation. Though this approach has poor security with XOR encryption. In [24] DNA and GA based hybrid approach is proposed for encryption of image. To generate pseudo-random sequence, Non-linear feedback shift register (NLFFSR) is utilized. In the crossover operation, this sequence is utilized for the image encryption. Due to the unpredictable nature of pseudo-random binary sequence of NLFFSR, this approach is safe. Though this approach is safe, it takes more time for the binary sequence calculation. Similarly, authors of [25], proposed a GA based approachfor the security of data. There are two stages in this approach. For the prevention of attacks, this approach can be applied for intrusion detection. In [26] a symmetric key linear algorithm is described that assures confidentiality of network. Because of the linear substitution, this approach is not secure. With the frequency analysis, linear substitution can be determined easily. GA based fast Arabic encryption method is presented in [27]. In this approach a plaintext and 8 bit binary static key is generated at first. Then the GA operations such as mutation and crossover is performed to generate the ciphertext. Though this approach is fast, it is applicable only for Arabic and static key is utilized for encryption. A symmetric security scheme based on DNA is proposed in [28] to handle the binary data in the form of DNA. Even though, randomness to the algorithm is provided by DNA, in real applications it is not applicable and slower. GA based encryption is utilized in several approaches.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

III. PROPOSED METHODOLOGY

3.1 Optimized Mondrian based k-anonymization

If the database contains at least k-1 individual rows in the database that shares commonidentifiers, then it said to be k-anonymized database. By using suppression and generalization techniques, the QI in the database can be made common. In order to group the attributes, the actual attribute with a semantically correct and less specific value is replaced in generalization method. In the suppression method actual attributes are replaced with a special character such as '*'. Mondrian is a multidimensional k-anonymity algorithm that follows a greedy approach. It partitiones the QI into different group of generalized attributes. The k-anonymity property would be satisfied by each partition. The greedy partitioning algorithm partitions the multidimensional regions that cover the domain space. The statistics are constructed based on the recoding functions from each region. The relaxed partitioning and strict partitioning algorithms are included in this algorithm. From the attributes, the median value is measured and partitioned in strict partitioning algorithm. For partitioning, the split values are provided manually in the relaxed partitioning algorithm. The figure 1 shows the block diagram of the proposed approach.

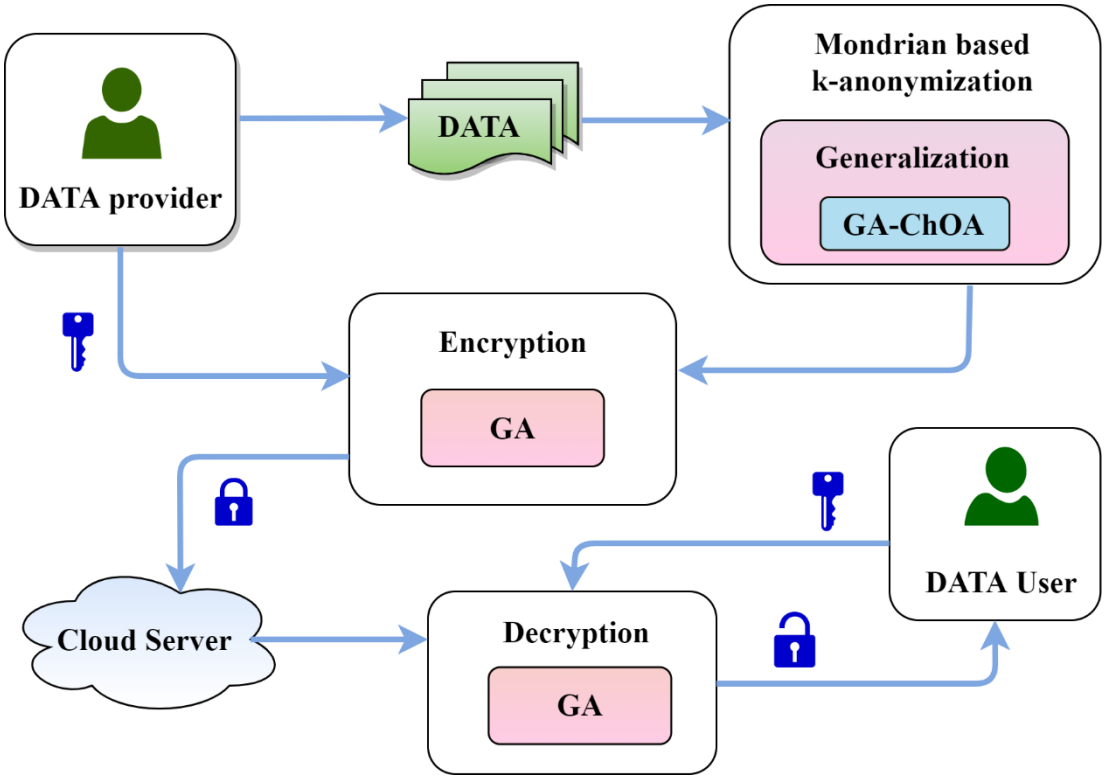


Figure 1. Block Diagram of proposed approach

In order to prevent the anonymized database from convergence to the local optimum and to attain better convergence rate, the hybrid model of genetic and chimp optimization. When compared to other optimization algorithms, GA-CHOA has cost effective performance and less convergence time. The two main steps involved in the GA-CHOA are as (1) Chimp optimization algorithm based optimization (2) Genetic algorithm based optimization

Driver, chaser, barrier and attackers are known as the the four types of chimps in a chimp clony. The prey is followed by driver. A dam is built by by the barriers by place themselves in a tree across the prey progression. To catch up with the prey, the chasers move rapidly. At last, the attackers hunt the prey. For a successful hunt, the different abilities of these chimps are necessary. The hunting process of chimps can be categorized into two main phases:

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

exploration and exploitation. Chasing, driving and blocking the prey are in exploration and attacking the prey is in exploitation. During the exploitation and exploration phases, the prey is hunted. The algorithm 1 shows pseudocode for the GA-ChmpOA based anonymization approach. The driving and chasing of the prey can be written in mathematical model as given in Equation (1) and Equation (2)

$$d = \left|c.x_{prey}(t) - m.x_{chimp}(t)\right| \quad (1)$$

$$x_{chimp}(t+1) = x_{prey}(t)\text{-a.d} \quad (2)$$

Where, the count of current iteration is indicated by t. c,m and a are the coefficient vectors , the prey postion vector is represented by x_{prey}. The vectors c, a and m are calculated by using the equation (3), (4) and (5),

$$c = 2.r_2 \quad (3)$$

$$a = 2.f.r_1\text{-f} \quad (4)$$

$$m = \text{chaotic value} \quad (5)$$

Here, through the iteration process f is reducing non-linearly from 2.5 to 0. The vectors r_1 and r_2 are the random vectors in the range between 0 and 1. The chaotic vector m is calculated based on various chaotic map so that this vector represents the effect of sexual motivation of chimps in the process of hunting. The participation of the chaser, barrier and driver chimps in the process of hunting is irregular and no information about the prey's optimum location. For mathematical model, it is assumed that the optimum location of the prey is informed to the chaser, driver and barrier by the first attacker. The best four solutions are saved and positions of other chimps are updated based on the location of best chimps. It can be given by the equations (6-13)

$$d_{Attacker} = |c_1 x_{Attacker} - m_1 x| \quad (6)$$

$$d_{Barrier} = |c_2 x_{Barrier} - m_2 x| \quad (7)$$

$$d_{Chaser} = |c_3 x_{Chaser} - m_3 x| \quad (8)$$

$$d_{Driver} = |c_4 x_{Driver} - m_4 x| \quad (9)$$

$$x_1 = x_{Attacker} - a_1(d_{Attacker}) \quad (10)$$

$$x_2 = x_{Barrier} - a_2(d_{Barrier}) \quad (11)$$

$$x_3 = x_{Chaser} - a_3(d_{Chaser}) \quad (12)$$

$$x_4 = x_{Driver} - a_4(d_{Driver}) \quad (13)$$

The optimum location of new chimps can be obtained using the equation (14).

$$x(t+1) = \frac{x_1+x_2+x_3+x_4}{4} \quad (14)$$

In the GA-CHO algorithm the value of x_5 is included. The GA algorithm is utilized to calculate the value of x_5 as given in equation (15).

$$x(t+1) = \frac{x_1+x_2+x_3+x_4+x_5}{5} \quad (15)$$

In order to produce offspring of the next generation, the fittest chromosomes are selected from population for reproduction. Intitial population, fitness function, selection, crossover and mutation are the five phases in the GA. The population is initilized with a set of solutions to the problem. Based on the fitness score the fitness of the each chromosome is determined and selected for reproduction. In the selection process based on the fitnesss score two chromosomes are selected. This step is essential for calculating x_5. The selected two best solutions 'a' and 'b' have the best fitness score than other chromosomes. In the crossover process, new offsprings are created and added to the population. The position of the chromosomes are given as in equation (16)

$$[\alpha_1, \alpha_2] = X_a \otimes X_b \quad (16)$$

Where, the positions of chromosomes 'a' and 'b' are represented as X_a and X_b. The offsprings are represented as α_1 and α_2. To maintain the population diversity, mutation process is performed. To generate two more

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

chromosomes, random numbers are utilized. The obtained chromosomes at the mutation process are given in equation (17),

$$[\beta_1, \beta_2] = \alpha_1 \circ \alpha_2 \quad (17)$$

The x_5 is selected based on the best position of the mutated chromosomes as given in Equation (18),

$$x_5 = \text{best } [\beta_1, \beta_2] \quad (18)$$

The best information is generated with higher degree of utility and privacy can be obtained from the optimization by the fitness value. The fitness value is calculated using average equivalence value (CAvg) and generalized loss of Information (GIloss). Higher degree of privacy and utility is used to evaluate the fitness. The privacy is determined with Cavgand utility is determined withGIloss.Though the privacy and utility is inversely propositional to Cavg and GIloss respectively, the minimum value of CavgandGIloss indicates that privacy and utility is maximum respectively.

Algorithm 1: Pseudo code for data anonymization using GA-CHOA
//Mondiran based K-anonymization
Partition the dataset into k-groups using kd-tree
Calculate fitness value
Determine QID using GA-CHOA
//Chimp optimization algorithm
Calculate X1, X2, X3, X4 using fitness value
//Genetic algorithm
Selection
Crossover
Mutation
Calculate X5 using fitness value
The best QI numberis calculated by calculating (X1+X2+X3+X4+X5)/5
return anonymized data

3.2 Data encryption using genetic-DNA algorithm

The proposed genetic-DNA encryption technique consist of encryption, decryption and random key generation. In the encryption process, the text is converted to ASCII and then to DNA code. The translated final cypher text is obtained from this initial cypher text by using the DNA sequences from random key generation. The first stage of encryption is preprocessing in which the corresponding ASCII values of all the characters are translated into 8 bit binary data. Adenine (A), Cytosine(C), Guanine (G) and Thymine (T) are the four bases present in DNA to which the two adjacent bits is transferred. In the proposed approach the symmetric key algorithm is utilized for encryption stage. After the conversion of binary data to DNA sequence, encryption is performed using the key. If DNA sequences of key or data is in binary form then XOR operation will be performed on the corresponding DNA sequences. Then result of XOR operation is converted to the DNA sequence.

After the encryption stage, replication, mutation and crossover are performed based on GA algorithm. To generate the genetic material that moves to next iteration and activity in the chromosome population form, the reshaping process is utilized. The first chromosome length and number are determined in this step. These values can be varied and constant for each round. To create the parent chromosomes with predefined length it is reshaped by aligning the DNA sequence into rows. After the construction of parent chromosomes, the crossover process is performed. In this process two crossover forms are utilized sequentially. The parents are chosen in the mating pool in the first form. Then two new offspring are produced by swapping the heads of parents. In the second form, a predefined value is utilized to rotate right or left. After the crossover process, the string elements are modified using mutation that has two forms. In the first form, the data is converted into binary vector and two mutation points are defined between the last and first bits. Then complement of bits are taken. This encrypted data is saved in the cloud. Then, crossover, mutation, decryption is performed to get the original file. By reversing the encryption operation decryption is performed. In the decryption

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

process cloud data is converted to DNA sequence and reshaped. The algorithm 2 shows the pseudo code for the DNA-GA based encryption approach.

Algorithm 2: Pseudo code DNA-GA based encryption

Data ←Read anonymized Input text
Data1 ← Binarize Data using ASCII code conversion
Reshape Data1
Data2 ← DNA bases
while (Round No not equal to Zero) **do**
//Encryption
Encrypt Data2 with key
//GA
Reshape
Crossover operation
Mutation
End while
Reshape
Save encrypted text file in the cloud
Data3 ← Binarize the encypted text
while (Round No not equal to Zero) **do**
//GA
Mutation
Crossover operation
Reshape
//Decryption
Decrypt Data3 with key
End while
Reshape
Save the file

IV. EXPERIMENTS

The proposed model is implemented in Google Colab.The pysparkcloud is utilized to store the encrypted database. In this section, the database description and evaluation metrics are discussed.

4.1 Database description

The adult database, Diabetes, Liver and Covid datasets are utilized for experimental evaluation of the proposed model.

(a) Adult Database

This is known as census income dataset. It is for predict if the income exceeds above 50K per year based on the census data. Extraction was done by Barry Becker from the 1994 Census database. This database is available at https://www.kaggle.com/uciml/indian-liver-patient-records.

(b) Indian Liver Patient Records

This database has records of 167 non-liver patientand 416 liver patient that was collected from North East of Andhra Pradesh, India. The "Dataset" column is a class label used to divide groups into liver patient (liver disease) or not (no disease). This databasehasrecords of 441 male patientand 142 female patient. This dataset is available at https://www.kaggle.com/uciml/indian-liver-patient-records.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

4.2 Evaluation metrics

Fitness value, Generalized Information Loss (GI_{Loss}) and Average equivalence value metric (C_{Avg}) are the evaluation metrics utilized to analyze the privacy and utility of the proposed model.

(a) Generalized Information Loss (GI_{Loss})

To obtain the minimum fitness value of populationGI_{Loss} is an efficient metric. It can be calculated using equation (19),

$$\mathbf{GI_{Loss}} = \frac{1}{R \times QI} \times \sum_{x=1}^{QI} \sum_{y=1}^{R} \frac{U_{xy} - L_{xy}}{U_x - L_x} \quad \textbf{(19)}$$

Here, the bounds of xth quasi identifier is represented by $U_x$and L_x.

(b) Average equivalence value metric (C_{Avg})

To obtain the minimum value to predict the target,C_{Avg} is a significant metric.It can be calculated as given in equation (20),

$$\mathbf{C_{Avg}} = \frac{R}{|SI_{q \times QI}| \times k} \quad \textbf{(20)}$$

Where, Sensitive attribute is indicated by SI.

(c) Fitness value (Fit)

To maintain the privacy it is calculated for each iteration. It can be calculated by using equation (21),

$$\mathbf{Fit} = \propto * \mathbf{GI_{Loss}}(A) + \beta * \mathbf{C_{Avg}}(A) \quad \textbf{(21)}$$

Where, $\propto$ and β are the constants.

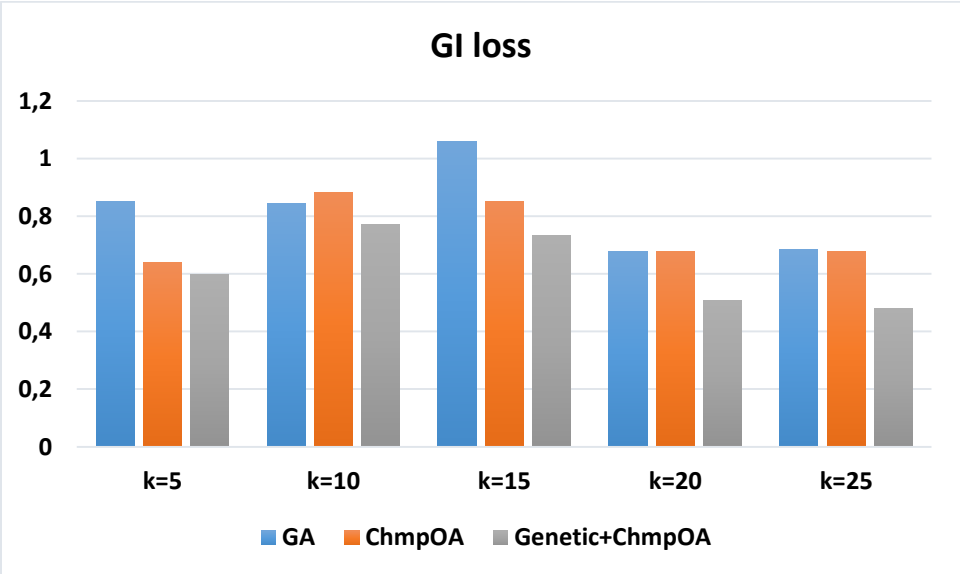
V. EXPERIMENTAL RESULTS

In this section experimental results for proposed data publishing model is detailed.

5.1 Performance analysis for adult database

For various levels of k-anonymization, GI loss is explained in Figure 2.For k=5 the GIloss of GA, ChmpOA and proposed Genetic ChmpOAare 0.8525, 0.6403 and 0.5985 respectively. For k=25 the GIloss of GA, ChmpOA and proposed Genetic ChmpOA are 0.6849, 0.6772 and 0.4811 respectively. From the Figure 2 it can be observed that the GIloss is higher for GA and lower for GA-ChmpOA. Though the proposed approach has the minimum GIloss it has the maximum degree of utility.

Figure 2. GI loss



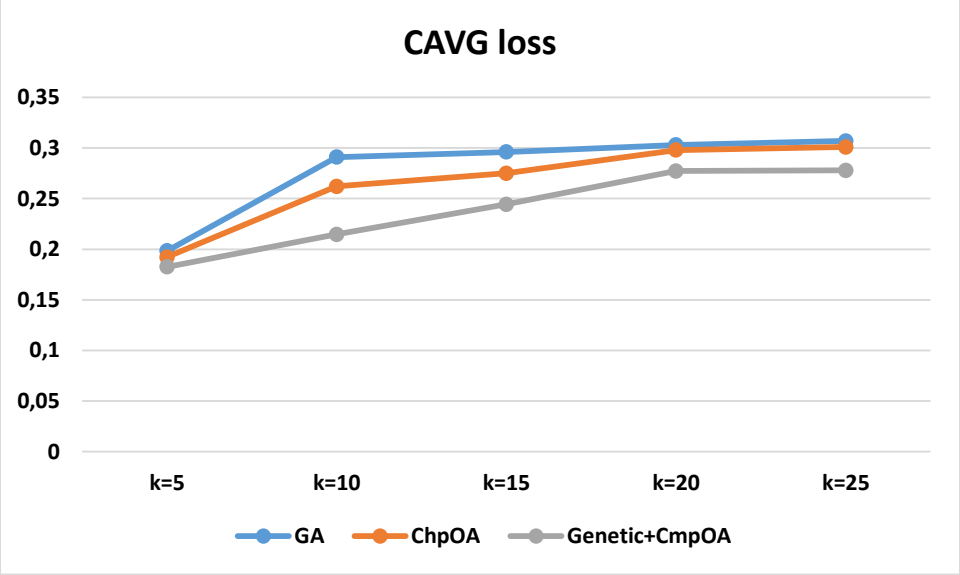
| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

For various levels of k-anonymization, Cavg is explained in Figure 3. For k=5 the Cavg of GA, ChmpOA and proposed Genetic ChmpOA are 0.1826, 0.0913 and 0.0365 respectively. For k=25 the GIloss of GA, ChmpOA and proposed Genetic ChmpOA are 0.2772, 0.1386 and 0.0110 respectively. From the Figure 3 it can be observed that the Cavgis higher for GA and lower for GA-ChmpOA. Though the Cavg is minimum for proposed approach, it has the higher degree of privacy.

Figure 3. CAVG loss



The Table 1 shows the running time taken by the proposed approach.The GA and ChmpOA algorithm takes less time when compared to proposed approach. Though the running time is higher than other approaches, the difference is in seconds. Table 2 shows the fitness value for the optimization based Mondrian k-anonymization.

Table 1. Running time

Algorithms	k=5	k=10	k=15	k=20	k=25
GA	1.36	0.88	0.82	0.72	0.71
ChpOA	1.39	0.90	0.79	0.75	0.77
Genetic+CmpOA	1.48	0.95	0.86	0.89	0.89

Table 2. Fitness value

Algorithms	k=5	k=10	k=15	k=20	k=25
GA	0.7962	0.6241	0.6967	0.6752	0.0625
ChpOA	0.2226	0.3030	0.3422	0.3859	0.3859
Genetic+CmpOA	0.0164	0.0186	0.0204	0.0625	0.0250

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

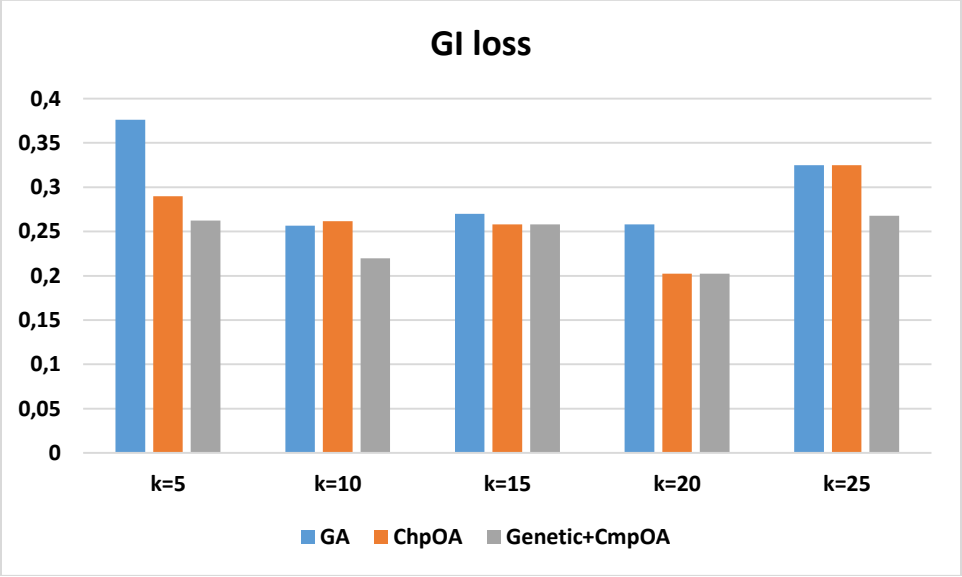
||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

5.2 Performance analysis for Indian Liver Patient Records

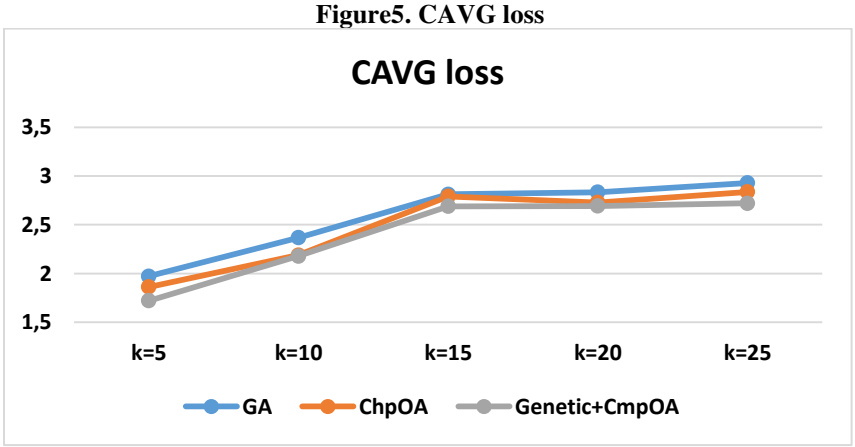
For various levels of k-anonymization, GI loss is explained in Figure 4. For k=5 the GIloss of GA, ChmpOA and proposed Genetic ChmpOA are 0.376, 0.29and 0.262 respectively. For k=25 the GIloss of GA, ChmpOA and proposed Genetic ChmpOA are 0.3249, 0.3249and 0.268respectively. From the Figure 4 it can be observed that the GIloss is higher for GA and lower for GA-ChmpOA. Though the proposed approach has the minimum GIloss it has the maximum degree of utility.

Figure4. GI loss



For various levels of k-anonymization, Cavg is explained in Figure 5. For k=5 the Cavg of GA, ChmpOA and proposed Genetic ChmpOA are 1.971, 1.862, 1.721. For k=25 the GIloss of GA, ChmpOA and proposed Genetic ChmpOA are 2.929, 2.838 and 2.721. From the Figure 5 it can be observed that the Cavgis higher for GA and lower for GA-ChmpOA. Though the Cavg is minimum for proposed approach, it has the higher degree of privacy.

Figure5. CAVG loss



The Table 3 shows the running time taken by the proposed approach. The GA and ChmpOA algorithm takes less time when compared to proposed approach. Though the running time is higher than other approaches, the difference is in seconds. Table 4 shows the fitness value for the optimization based Mondrian k-anonymization.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

Table 3. Running time (Total)

Algorithms	k=5	k=10	k=15	k=20	k=25
GA	0.769	0.789	1.484	1	0.749
ChpOA	1.95	1.009	1.1	0.84	1.26
Genetic+CmpOA	0.88	1.69	0.96	1.14	1.146

Table 4. Fitness value

Algorithms	k-level				
	k=5	k=10	k=15	k=20	k=25
GA	0.0548	0.0108	0.066	0.01340	3.605
ChpOA	0.0833	0.0945	0.1081	0.1081	3.7396
Genetic+CmpOA	0.0085	0.0108	0.1081	0.0134	3.7396

VI. CONCLUSION

To provide the security and preserve the privacy of the medical databases, cost effective technique is proposed in this paper.AnefficientMondrian based k-anonymization technique with GA-ChmpOAis utilized toanonymize the database. For the additional security of the anonymized data, Genetic-DNA based encryption technique is utilized.The proposed approach shows an outstanding performance when compared to existing approaches in terms of privacy and utility. In the proposed approach, the information loss is minimized with the higher values of privacy and utility. In future, various datasets and more optimization algorithms can be utilized to extend the proposed approach.

REFERENCES

1. Zhan, Z.H., Liu, X.F., Gong, Y.J., Zhang, J., Chung, H.S.H., Li, Y.: Cloud computing resource scheduling and a survey of its evolutionary approaches. ACM Comput. Surv. 47(4), 63 (2015)
2. Kundalwal, M. K., Chatterjee, K., & Singh, A. (2019). An improved privacy preservation technique in health-cloud. ICT Express, 5(3), 167–172.
3. Sehgal, N. K., & Bhatt, P. C. P. (2018). Future trends in cloud computing. Cloud Computing. https :// doi.org/10.1007/978-3-319-77839 -6_12.
4. Romanou, A. (2018). The necessity of the implementation of privacy by design in sectors where data protection concerns arise. Computer Law and Security Review, 34(1), 99–110.
5. Arava, K., &Lingamgunta, S. (2019). Adaptive k-Anonymity Approach for Privacy Preserving in Cloud. Arabian Journal for Science and Engineering. doi:10.1007/s13369-019-03999-0.
6. J. Andrew, J. Karthikeyan and J. Jebastin, "Privacy Preserving Big Data Publication On Cloud Using Mondrian Anonymization Techniques and Deep Neural Networks," *2019* 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 722-727, doi: 10.1109/ICACCS.2019.8728384.
7. Ashkouti, F., khamforoosh, K., &Sheikhahmadi, A. (2020). DI-Mondrian: Distributed Improved Mondrian for Satisfaction of the L-diversity Privacy Model Using Apache Spark. Information Sciences. doi:10.1016/j.ins.2020.07.066.
8. McCall, J.: Genetic algorithms for modelling and optimisation. J. Comput. Appl. Math. 184(1), 205–222 (2005)
9. Khishe, M., &Mosavi, M. R. (2020). Chimp Optimization Algorithm. Expert Systems with Applications, 113338. doi:10.1016/j.eswa.2020.113338
10. Tahir, M., Sardaraz, M., Mehmood, Z., & Muhammad, S. (2020). CryptoGA: a cryptosystem based on genetic algorithm for cloud data security. Cluster Computing. doi:10.1007/s10586-020-03157-4.
11. Mousa, Hamdy. (2016). DNA-Genetic Encryption Technique. International Journal of Computer Network and Information Security. 8. 1-9. 10.5815/ijcnis.2016.07.01.
12. Adult Database, https://www.kaggle.com/uciml/indian-liver-patient-records.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107054 |

13. Indian Liver Patient Records, https://www.kaggle.com/uciml/indian-liver-patient-records.
14. Early Stage Diabetes Risk Prediction Dataset, https://www.kaggle.com/ishandutta/early-stage-diabetes-risk-prediction-dataset
15. COVID-19 patient pre-condition dataset, https://www.kaggle.com/tanmoyx/covid19-patient-precondition-dataset.
16. Shah, M.A., Swaminathan, R., Baker, M., et al.: Privacy-preserving audit and extraction of digital contents. IACR Cryptol. ePrint Arch. 2008, 186 (2008)
17. Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), 571– 588. https ://doi.org/10.1142/s0218 48850 20016 5x.
18. S. Kim and Y. D. Chung, "An anonymization protocol for continuous and dynamic privacy-preserving data collection," Futur. Gener. Comput. Syst., vol. 93, pp. 1065–1073, Apr. 2019.
19. R. Elgendy, A. Morad, H. G. Elmongui, A. Khalafallah, and M. S. Abougabal, "Role-task conditional-purpose policy model for privacy preserving data publishing," Alexandria Eng. J., vol. 56, no. 4, pp. 459–468, Dec. 2017.
20. K. LeFevre, D.J. DeWitt, R. Ramakrishnan, Mondrian multidimensional K-anonymity, Proc. – Int. Conf. Data Eng. 2006 (2006) 25.
21. S. Yaseen et al, Improved generalization for secure data publishing, IEEE Access 6 (2018) 27156–27165.
22. BibalBenifa, J. V., &Venifa Mini, G. (2020). Privacy Based Data Publishing Model for Cloud Computing Environment. Wireless Personal Communications. doi:10.1007/s11277-020-07320-3.
23. Dalimunthe, A.R.: Modifikasivernam cipher denganpengoptimalankuncimenggunakan genetic algorithm (2018).
24. Pujari, S.K., Bhattacharjee, G., Bhoi, S.: A hybridized model for image encryption through genetic algorithm and dna sequence. Proc. Comput. Sci. 125, 165–171 (2018).
25. Ijaz, S., Hashmi, F.A., Asghar, S., Alam, M.: Vector based genetic algorithm to optimize predictive analysis in network security. Appl. Intell. 48(5), 1086–1096 (2018).
26. Sindhuja, K., Devi, S.P.: A symmetric key encryption technique using genetic algorithm. Int. J. Comput. Sci. Inf. Technol. 5(1), 414–416 (2014).
27. Abduljabbar, R.B.: Fast approach for arabic text encryption using genetic algorithm. Eur. J. Sci. Res. 144(4), 342–348 (2017)
28. Amin, S.T., Saeb, M., El-Gindi, S.: A DNA-based implementation of YAEA encryption algorithm. In: Computational Intelligence, pp. 120–125 (2006).

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details