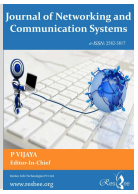
Vol.7 No.4 Oct 2024

Journal of Networking and Communication Systems

Received 9 August, Revised 13 August, Accepted 3 October



A Novel Privacy-Preserved Iot Data Protection Model Using APO for Optimal Key Generation

B.Priyanka

Assistant Professor, Department of Electronics and Communication Engineering
J. P. College of Engineering, Agarakattu, Tamil Nadu
Priyankadharshini7@gmail.com

Abstract: The Internet of Things (IoT) has paved the way for a highly interconnected world, where devices exchange information seamlessly, enhancing the quality of human life across various applications. As the number of connected devices continues to rise, concerns about sensitive data disclosure have emerged as a critical issue. Users' privacy on IoT devices is particularly concerning due to the threats posed by malware and hackers, especially given the rapid expansion of IoT in commercial and critical infrastructure settings. To address these challenges, this research introduces an effective approach for a privacy-preserved IoT data protection model called Artificial Protozoa Optimization (APO) for optimal key generation. The data owner uploads information to the cloud, converting it to polynomial values for privacy. RV coefficients are identified to assess relationships, and bilinear transformation anonymizes data. An optimal key is generated and encrypted, allowing users to access the original data. This approach aims to enhance the security and privacy of IoT data, making it a valuable contribution to the field. In this process, the data owner first uploads all information to a cloud model. The established APO method has demonstrated impressive results, achieving an overall conditional privacy rate of 93.456% and a normalized variance of 0.921.

Keywords: Internet of Things, Privacy, Security, Privacy Preservation, Data Protection

Nomenclature

Abbreviation	Expansion
APO	Artificial Protozoa Optimization
PII	Personally Identifiable Information
IoE	Internet of Everything
IIoT	Industrial Internet of Things
NBPPM	Noise-Based Privacy-Preserving Model
G–BHO	Grasshopper–Black Hole Optimization
AI	Artificial Intelligence
HEDS4IoT	High-performance Evolutionary Data Sanitization for IoT
ECC	Elliptic Curve Cryptography

1. Introduction

IoT devices are integral to our daily experiences, driving the technological transformation alongside wireless technologies and sophisticated communication systems. As a fundamental element of the digital age characterized by Industry 4.0, IoT has gained prominence. The integration of physical objects into the digital landscape is becoming increasingly achievable due to technological advancements [9]. IoT networks impact numerous aspects of everyday life, such as home security and healthcare monitoring. By utilizing data processing and analytics, IoT harnesses the internet's capabilities to enhance decision-making for tangible objects in the real world. This system connects intelligent objects, enabling them to access the internet and form a network that collects and exchanges information through "Things." As a result, IoT has emerged as a key area of research on a global scale [6]. Its applications span a wide array of fields, including smart cities, advanced healthcare systems, intelligent buildings, efficient transportation, eco-friendly environments, industrial operations, agriculture, supply chain optimization, innovative retail, and

https://doi.org/10.46253/jnacs.v7i4.a1

location-based services. These applications often handle sensitive data, such as health information, financial records, location data, PII [6], and various aspects of personal life [3]. As a result, privacy and security are critical concerns [8]. In the context of IoT, specific issues regarding security have been raised. Secure communication is vital for maintaining data privacy and confidentiality across IoT and IoE architectures [6].

In networks based on IoT, smart devices collect personal data, and a lack of adequate privacy measures can lead to the improper use of sensitive information. If such personal information is compromised, the consequences can be severe [14] [3]. Network security is significantly impacted by both external and internal intruders [7]. While providing external security is relatively straightforward, ensuring internal security presents a considerable challenge. Internal intruders often masquerade as legitimate users within the network. They tend to exploit authorized nodes to compromise the entire cloud network by targeting sensitive information [10]. External security employs advanced firewall software to safeguard the system against outside threats. However, compared to external threats, insider attacks are considerably more perilous. Insider attacks typically occur by exploiting vulnerabilities, such as utilizing an authorized node, launching attacks with an authorized ID, masquerading as a trusted node, or stealing sensitive information while posing as a legitimate user [7]. Privacy is a fundamental right of individuals that allows them to keep their information confidential and maintain control over it [11]. Preserving privacy is a crucial factor that must be addressed in all existing physical and logical systems to reduce the risk of privacy breaches. Ensuring information privacy has become an increasing concern for governments, businesses, consumers, and others alike [9] [3]. Privacy preservation in data mining is an area of research that is currently gaining traction. Privacy preservation essentially involves obtaining results from data mining algorithms without compromising the underlying dataset [8].

This paper presents an effective approach for privacy preserved IoT data protection model using APO for optimal key generation. Initially, the data owner uploads all information to the cloud model. To ensure privacy, the input data is converted into polynomial values. Next, the RV coefficient is identified between the features and the class labels in the input data. Bilinear transformation is then applied to anonymize the input data, thereby enhancing privacy. The optimal key is generated using the APO technique, followed by encryption to protect the data. Finally, the user who requires the data accesses the encrypted information from the cloud and retrieves the exact data.

The contribution of the paper is outlined as follows:

- ***APO for* privacy preservation in IoT*:*** In this work, data privacy is ensured in the IoT network by utilizing multiple operations, such as polynomial generation, binomial transformation, and data encryption. Here, data encryption is done by using the APO for optimal secret key generation.

The remainder of this article is structured as follows: Section 2 provides the literature review of existing techniques for privacy preservation in IoT. Section 3 outlines the system model for privacy preservation in IoT. Section 4 presents the proposed method for privacy preservation. Section 5 provides the results and discussions of the developed approach. At last, Section 6 concludes the paper with a discussion on future work.

2. Motivation

This section presents a literature review of existing approaches for privacy preservation in IoT. Various experts have proposed methods for privacy preservation in IoT; however, many of these methods yield unsatisfactory results. To address this issue, a new technique for a privacy-preserved IoT data protection model called APO is developed. The advantages as well as the drawbacks of the current methods are discussed below.

2.1 Literature Survey

Wang, W., *et al.* [1] developed a Token-Based Access Control System aimed at decentralized access control in the IIoT. This system boasts strong security, reliability, and scalability features. However, it lacked practicality and convenience and did not facilitate the migration of the planned access control architecture across different blockchain networks. El-Gendy, S., *et al.* [2] anticipated an AdaBoost approach to address confidentiality and privacy challenges in IoT systems. This algorithm effectively detected and predicted both known and unknown malware, achieving a high F-score. Nevertheless, it did not tackle other privacy concerns in IoT and was not been tested in various real-world use cases or evaluated against live attacks. Jain, S.K. and Kesswani, N., [3] introduced a NBPPM designed to protect sensitive data while ensuring effective privacy preservation with relatively low computational costs and reduced execution time. However, this method lacked appropriate accountability procedures, limiting control over personal information in IoT environments. Kumar, M., *et al.* [4] presented the G–BHO algorithm to implement a privacy preservation model in IIoT, leveraging advancements in AI. While this method effectively

addressed real-world optimization issues, it did not apply to other IoT applications where privacy preservation was critical. Telikani, A., *et al.* [5] developed the HEDS4IoT method to enable real-time streaming of protected data and enhance the efficiency of solution quality assessments. Despite its lower computational time and improved efficiency, this technique has not been implemented in conjunction with other big data processing mechanisms.

2.2 Challenges

The challenges faced by the existing optimal secret key generation methods are detailed below:

- ✓ In [5], the HEDS4IoT technique reduced side effects and achieved average performance improvements. However, this technique required significant execution time.
- ✓ The AdaBoost technique in [2] maintained data privacy while reducing the impact of noisy data. Nevertheless, it failed to consider the potential trade-off between accuracy and decentralization, which is crucial for achieving optimal performance, even in scenarios where fog computing enhances IoT.
- ✓ The Token-Based Access Control System method in [1] reduced a significant amount of gas compared to expectations. However, this approach was not suitable for real-time scenarios.
- ✓ Challenges arose in optimal secret key generation because recent IoT systems fell short of meeting the necessary functional requirements and addressing security and privacy risks. Moreover, the existing schemes fail to provide better privacy protection.

3. System Model

The model [13] includes three key entities: cloud users, data owners, and cloud servers. A cloud user is someone who accesses the information stored in the cloud. When a cloud user initiates a service request, the query mapper processes this request through the request handler, which matches it with the appropriate query in the database to grant access to the necessary information. The cloud server comprises various management servers, including both virtual and physical servers. To protect data security, data owners implement effective privacy protection techniques to safeguard their information before sharing it with cloud users or third parties, which can include organizations or agents needing access for business purposes. However, no actual or sensitive information about individuals must be disclosed. Transforming actual data into a protected format while concealing sensitive information poses a significant challenge. Additionally, maintaining the utility of the data is vital to ensure that the information provided to cloud users remains effective, even when protecting sensitive content. In this context, data owners create a key to retrieve the original data from its protected format when necessary. Fig. 1 illustrates the cloud model designed for the privacy protection of this data.

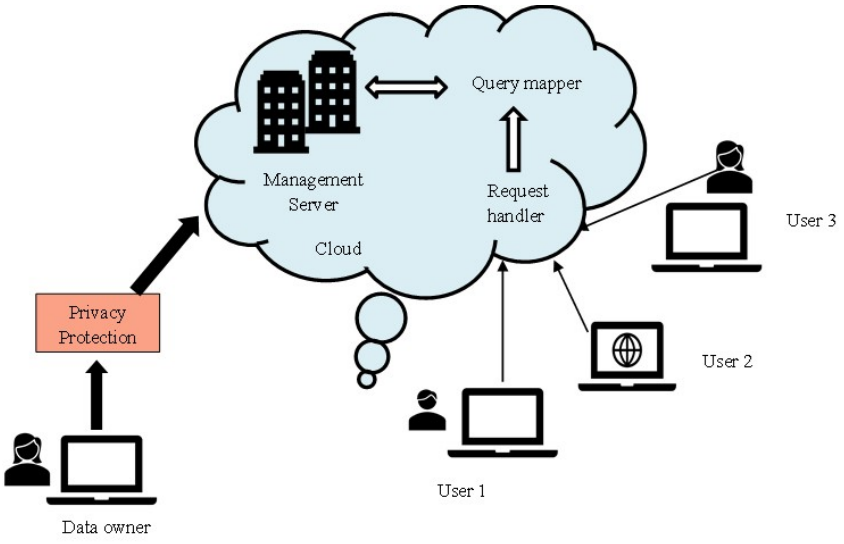


Fig. 1. *Cloud model utilized for the privacy protection of the data*

4. Proposed Model

The primary goal of this research is to establish a privacy-preserved IoT data protection model using APO for optimal key generation. Initially, the data owner uploads all information to the cloud model. To ensure

privacy, the input data is converted into polynomial values. Next, the RV coefficient is identified between the features and the class labels in the input data, to determine the relationship between the data and privacy. Bilinear transformation is then applied to anonymize the input data, thereby enhancing privacy. The optimal key is generated using the APO technique, followed by encryption to protect the data. Finally, the user who requires the data accesses the encrypted information from the cloud and retrieves the exact data. Fig. 2 illustrates the process of data privacy protection with a secret key employing the APO approach.

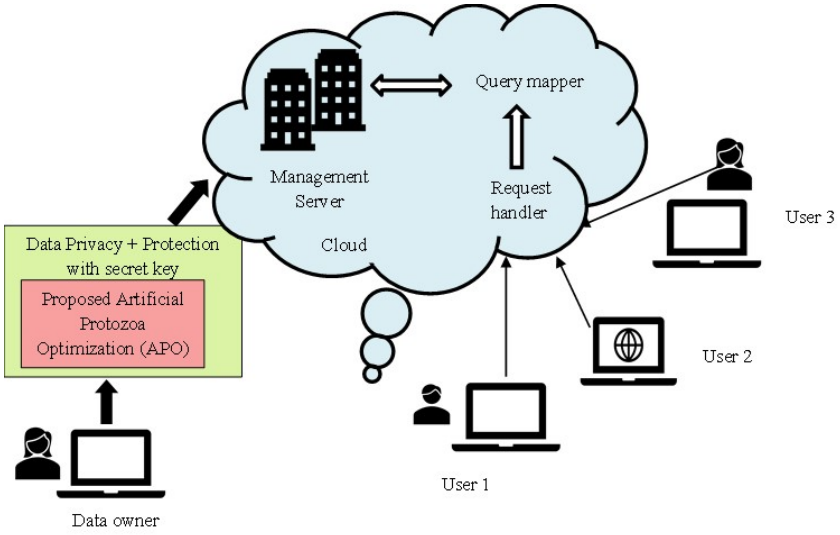


Fig. 2. *Data privacy protection with secret key using APO*

4.1 Privacy Preserved Data Model

Assume that the data owner needs to upload the data to the cloud. Ahead of uploading the data, it is subjected to a privacy protection model to ensure data privacy. Let the data A with dimensions $V \times S$ be considered as input. The steps involved in creating a privacy-preserved data model are as follows,

Step-1 Polynomial Value

To ensure privacy, the input data is first converted into a polynomial value. Here, Chebyshev polynomials are used as they are effective in obscuring sensitive information Also, their high efficiency enables them to handle high dimensional data. The input data is transformed into a polynomial value as expressed below,

$$D = 36y^6 + 4y^5 + 2y^3 + 16 \tag{1}$$

where,

$$y = \sum_{p=1}^{V} \sum_{q=1}^{S} A_{pq} \tag{2}$$

Here, D represents the Chebyshev polynomial, the coefficient or constants are denoted as y, p represents the p^{th} row of the data matrix, and q denotes the q^{th} column of the data matrix.

Step-2 RV coefficient between features i and class label j

The next process is to find the RV coefficient among the feature and the class labels in the input data. This process assists in determining the association between the input data and privacy. The RV coefficient [19] or RV correlation coefficient, measures the association between two sets of variables. When considering the features i and class labels j, the RV coefficient quantifies the relationship between the variability of these features and class labels.

$$V = RV(i, j) \tag{3}$$

$$RV(i, j) = \frac{COV(i, j)}{\sqrt{VAR(i), VAR(j)}} \tag{4}$$

Where, V represents the parameter in which $RV(i, j)$ is stored, i demonstrates the feature, and j denotes the class label, the covariance matrix between i and class label j is denoted as $COV(i, j)$, $VAR(i)$ denotes the covariance matrix of the feature, and $VAR(j)$ denotes the covariance matrix of the class label.

Step-3 Store the V value in the parameter B

The RV coefficient thus attained is a $1\times S$ vector, and hence, it has to be transformed into a 1×1 matrix for it to have the same dimension as the polynomial value. Hence, the value V is stored in the parameter B as specified below.

$$B = V_{(1\times S)} * V^{T}_{(S\times 1)} \tag{5}$$

Here, S denotes the dimension of the data.

Further, the polynomial and RV coefficient values are stored in a parameter G as depicted below.

$$G = D_{(1\times 1)} * B_{(1\times 1)} \tag{6}$$

Step-4 Bilinear matrix

Bilinear transformation is applied to anonymize the input data thus enhancing privacy. A bilinear matrix is a mathematical structure that encompasses two vector spaces and a bilinear mapping. Simply put, it represents the interactions between two sets of variables linearly, allowing the output to be expressed as a combination of inputs from both sets. A bilinear matrix refers to a matrix that arises in the context of bilinear forms, which are mathematical expressions that involve two vectors.

$$H = A_{V\times S} * G \tag{7}$$

Where the bilinear matrix is signified as H.

Step-5 Concatenation process

In this process, the optimal key generated using the APO technique is concatenated with the value G. This is effectuated to add a layer of security to the data as only an authorized person can construct the key. Concatenation is done as follows,

$$F = Y \,||\, G \tag{8}$$

Here, Y the optimal secret key is generated utilizing APO and F is a parameter for storing the secret key and G value.

Step-6 Privacy preserved data based on Encryption

Privacy preservation involves techniques and methods used for protecting sensitive information while enabling its use in various applications. This approach ensures that the data remains confidential, accessible, and secure to authorized users. Encryption is performed using the ECC to ensure the privacy of the data as follows,

$$C = E_c\left(H, F\right) \tag{9}$$

Where, E_c symbolizes the encryption function, and C signifies the encrypted data obtained from the encryption process. The encrypted data is stored in the cloud.

Step-7 Data retrieval

Data retrieval is the process of obtaining data from a database, storage system, or other data sources. It typically involves querying the data using specific parameters to extract the desired information. Here, the user who needs the data accesses the encrypted data C from the cloud and retrieves the original data as follows,

$$H^{*} = D_e\left(C, F\right) \tag{10}$$

$$A^{*} = \frac{H^{*}}{F} \tag{11}$$

Here, the decryption function is denoted as D_e, the inverse of the bilinear matrix is signified as H^{*}, and A^{*} denotes the retrieved data.

4.2 Secret key generation using APO

APO [16] is a nature-inspired algorithm that simulates the foraging behavior of protozoa to generate secure cryptographic keys. In this approach, protozoa act as potential solutions within a multi-dimensional search space, exploring it through random movement and adaptive learning. This method enhances security by leveraging biological principles, offering a novel and effective way to produce cryptographic keys with unpredictability and high entropy. The Solution encoding, fitness function, and APO are explained below.

4.2.1 Solution Encoding

Solution encoding typically refers to the method of transforming or representing the key material to ensure both efficiency and security. The retrieved data A^{*} is the output. Fig. 3 denotes the solution encoding.

Fig. 3. *Solution encoding*

Here, $1\times z$ signifies the size of the key.

4.2.2 Fitness Function

The fitness function is crucial for evaluating the quality of generated keys during secret key generation using APO. The fitness function is given as follows,

$$FF = \frac{1}{2}\left[CP + AC\right] \tag{12}$$

Where, the fitness function is signified as FF, the Conditional privacy [17] is denoted as CP, and AC is denoted as Normalized variance [18].

4.2.3 Artificial Protozoa Optimization

APO is an innovative computational technique inspired by the behavior of protozoa in their natural environments. This optimization algorithm is used to tackle complex problems, including optimal key generation in cryptographic systems. APO explores the solution space more effectively than traditional methods, improving the search for optimal cryptographic keys. Its capability to navigate multidimensional spaces makes it particularly effective in generating keys that balance security and efficiency. As cryptographic needs evolve, APO presents a promising approach to address the challenges of key generation, ensuring enhanced security in digital communications while adapting to diverse environmental conditions and threat landscapes.

The steps taken by APO are outlined below.

1) Initialization: The investigation of different biological phenomena reveals the distinctive advantages offered by microorganisms. Algae, protozoa, and bacteria perform roles akin to the organs found in higher plants and animals, utilizing specialized structures known as "organelles." These microorganisms display crucial life characteristics, including continuity, reproduction, genetic metabolism, the ability to adapt to environmental stimuli, and variability. Their simpler organization and reduced complexity often make microorganisms more efficient for various applications compared to higher organisms.

2) Fitness estimation: Fitness estimation is carried out using Equation (12) to identify the optimal solution, with the solution exhibiting the lowest fitness considered the ideal one.

3) Solution update: In the APO [16] method, the protozoa signify the solution set, with each protozoan having a position defined by *fkl* variables.

In foraging behavior, both external and internal factors of the protozoa are taken into account. Internal factors refer to the foraging characteristics of the protozoa, while external factors encompass environmental influences, such as competitive behaviors and species collisions. In Autotrophic mode, a protozoan can produce carbohydrates through chloroplasts to provide nutrition. When exposed to strong light intensity, it will relocate away from its current position towards an area with lower light intensity. The mathematical expression for Autotrophic mode is expressed as,

$$\mathrm{T}_k^{new} = \mathrm{T}_k + d\cdot\left(\mathrm{T}_a - \mathrm{T}_k + \frac{1}{zw}\cdot\sum_{\mu=1}^{zw}\omega_e\cdot\left(\mathrm{T}_{\mu-} - \mathrm{T}_{\mu+}\right)\right)\odot P_d \tag{13}$$

$$\mathrm{T}_k = \left[t_k^1, t_k^2, \ldots, t_k^{fkl}\right],\quad \mathrm{T}_k = sort\left(\mathrm{T}_k\right) \tag{14}$$

Here, T_k and T_k^{new} represent the original position and the updated position of the k^{th} protozoa respectively. The randomly selected a^{th} protozoa are denoted as T_a. $\mathrm{T}_{\mu-}$ Signifies a chosen protozoa in the μ^{th} neighbor pair with a rank lower than k. If T_k is T_1, $\mathrm{T}_{\mu-}$ is also set as T_1, $\mathrm{T}_{\mu+}$ symbolizes the randomly selected protozoan in the μ^{th} neighbor pair with a rank greater than k. P symbolizes the vector of mapping, zw indicates the number of paired neighbors, and the weight factor is denoted as ω_e.

In a heterotrophic mode, a protozoan can obtain nutrients by absorbing organic matter from its environment in the absence of light. When a nearby location represented by T_{near}, is rich in food, the protozoan moves toward it. The following mathematical model for this heterotrophic mode is expressed as,

$$\mathrm{T}_k^{new} = \mathrm{T}_k + d\cdot\left(\mathrm{T}_{near} - \mathrm{T}_k + \frac{1}{zw}\cdot\sum_{\mu=1}^{zw}\omega_x\cdot\left(\mathrm{T}_{k-\mu} - \mathrm{T}_{k+\mu}\right)\right)\odot P_d \tag{15}$$

$$\mathrm{T}_{near} = \left(1 \pm Rand \cdot \left(1 - \frac{itr}{itr_{\max}}\right)\right) \odot \mathrm{T}_k \tag{16}$$

In the context, Nearby location is denoted as T_{near}, "$\pm$" implies they T_{near} can be in different directions from the k^{th} protozoa respectively. $\mathrm{T}_{k-\mu}$ denotes the $(k-\mu)^{th}$ protozoa chosen from the μ^{th} neighbor pair, with its index rank as $k-\mu$. If T_k is T_1, then $\mathrm{T}_{k-\mu}$ is also defined as T_1. $\mathrm{T}_{k+\mu}$ symbolizes the $(k+\mu)^{th}$ chosen protozoan in the μ^{th} neighbor pair whose index rank is $k+\mu$. The maximum and the existing iteration are denoted as $itr_{\max}$ as well as itr.

In Dormancy, in response to environmental stress, a protozoan can enter a dormant state as a means of survival in unfavorable conditions. During this dormant phase, it is supplanted by a newly generated protozoan, ensuring the population remains stable. The mathematical model that describes this dormancy is represented as:

$$\mathrm{T}_k^{new} = \mathrm{T}_{\min} + Rand \odot \left(\mathrm{T}_{\max} - \mathrm{T}_{\min}\right) \tag{17}$$

$$\mathrm{T}_{\min} = \left[lb_1^{'}, lb_2^{'}, \ldots, lb_{fkl}^{'}\right], \quad \mathrm{T}_{\max} = \left[ub_1^{'}, ub_2^{'}, \ldots, ub_{fkl}^{'}\right] \tag{18}$$

Here, T_{mim} and $\mathrm{T}_{\max}$ denoted the upper and lower bound vectors respectively, $ub_{fl}^{'}$ and $lb_{fl}^{'}$ signifies the upper as well as lower bounds of the fk^{th} variables, respectively.

In reproduction, when they reach suitable health and age conditions, protozoa reproduce asexually through a process called binary fission. This process theoretically allows the protozoan to divide into two identical daughter cells. This behavior is modeled by generating a duplicate protozoan and applying a perturbation. The mathematical model for this reproduction is expressed as:

$$\mathrm{T}_k^{new} = \mathrm{T}_i \pm rand \cdot \left(\mathrm{T}_{\min} + Rand \odot \left(\mathrm{T}_{\max} - \mathrm{T}_{\min}\right)\right) \odot P_s \tag{19}$$

Here, "$\pm$" implies the perturbation can be reversed and forward. The vector for mapping in the process of reproduction is symbolized as P_s.

To integrate all the mathematical models, the relevant parameters are as follows:

$$pr = pr_{\max}.rand \tag{20}$$

$$p_{ut} = \frac{1}{2} \cdot \left(1 + \cos\left(\frac{iter}{iter_{\max}} \cdot \pi\right)\right) \tag{21}$$

$$p_{or} = \frac{1}{2} \cdot \left(1 + \cos\left(\left(1 - \frac{k}{wb}\right) \cdot \pi\right)\right) \tag{22}$$

In this context, the fraction proportion of reproduction and dormancy in the population of protozoa is denoted as pr, pr is the maximum value in $pr_{\max}$, Heterotrophic and autotrophic behavior probabilities are denoted as p_{or} and p_{ut}, and the size of the population is denoted as wb.

4) Feasibility evaluation: The feasibility of APO is assessed using the fitness function from Equation (12). The fitness of the designed solution is evaluated, and the result with minimal fitness is recognized as the best outcome.

5) Termination: The aforementioned stages are repeated until the desired outcome is reached. The iteration ends once the minimal fitness value is achieved.

5. Result and Discussion

The introduced APO approach for optimal key generation is validated by comparing its effectiveness with other optimization algorithms and existing methods. Additionally, the following sections provide detailed information on the experimental setup of APO, dataset description, evaluation metrics, data visualization, and comparative analysis are elaborated in further sections.

5.1 Experimental Setup

The developed APO approach for optimal key generation is implemented using the Python tool.

5.2 Dataset Description

The introduced APO for optimal key generation utilizes the heart disease dataset [15], which dates back to 1988 and comprises four databases: Cleveland, Hungary, Switzerland, and Long Beach. This dataset

includes 76 attributes, including the predicted attribute; however, all published experiments focus on a subset of 14. The "target" field indicates the presence of heart disease in the patient, represented as an integer, 1 = disease and 0 = no disease.

5.3 Evaluation Metrics

The introduced APO for optimal key generation is assessed for its efficiency utilizing metrics such as conditional privacy, and normalized privacy. The evaluation metrics are detailed below,

1) Normalized Variance: Normalized Variance [18] refers to the variance of the dataset that has been adjusted to a certain scale, often by dividing by the mean or some other value.

$$N = \frac{V_a}{M_e{}^2} \tag{23}$$

Here, N is the Normalized Variance, V_a is denoted as variance, and the mean is symbolized as M_e.

2) Conditional Privacy: Conditional Privacy [12] is based on conditional entropy and measures the privacy inherent in a random variable P, given a random variable Q.

$$CP = 2^{T(P|Q)} \tag{24}$$

In this context, the entropy is denoted as T, and the conditional privacy is signified as CP.

5.4 Data Visualization

The data visualization outputs from the Cleveland, Hungarian, and Switzerland datasets are illustrated in Fig. 4. In Fig. 4 (a), the data visualization of the Cleveland dataset is shown, focusing on the Chol attribute, which has feature values ranging from 200 to 350, with some values exceeding 350. The maximum recorded value for this attribute is over 500. Fig. 4 (b) depicts data visualization of the Hungarian dataset, also highlighting the Chol attribute, with feature values between 100 and 500, and a few values above 500. The maximum value for this attribute exceeds 600. In Fig. 4 (c), the data visualization of the Switzerland dataset is presented. The trestbps attribute has feature values ranging from 75 to 175, with some values exceeding 175. The maximum value for this attribute is above 200.

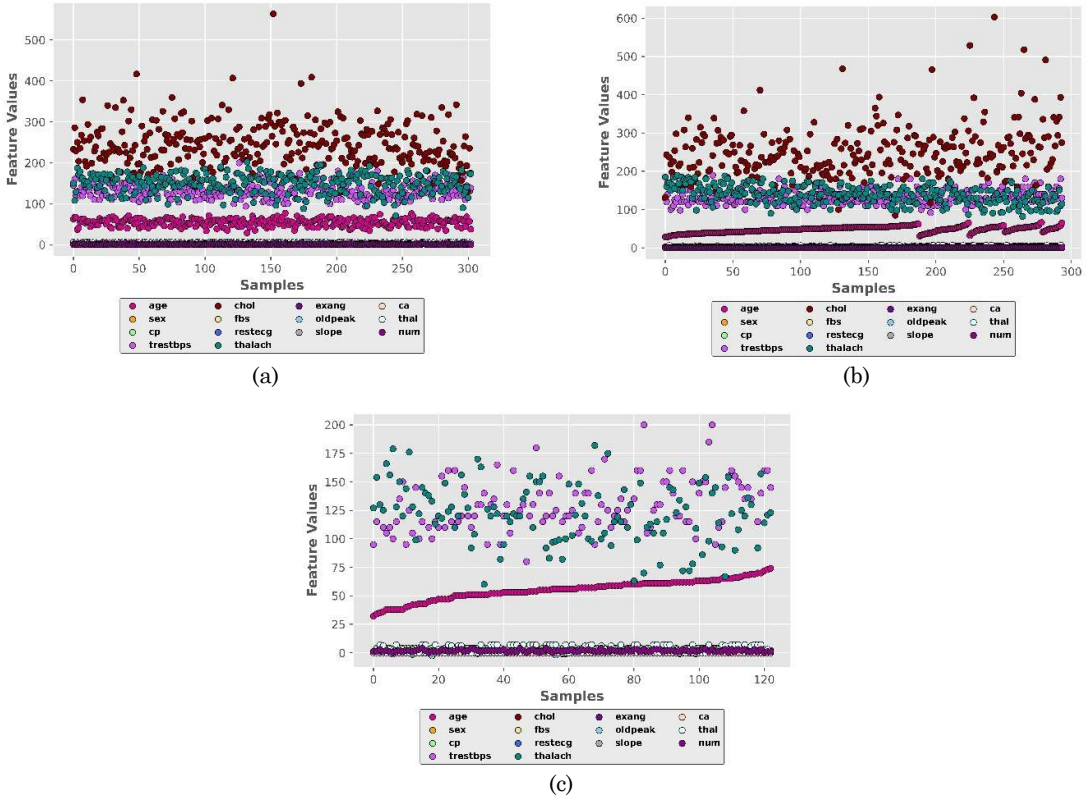


Fig. 4. *Data visualization of APO using (a) Cleveland (b) Hungarian (c) Switzerland*

5.5 Comparative Techniques

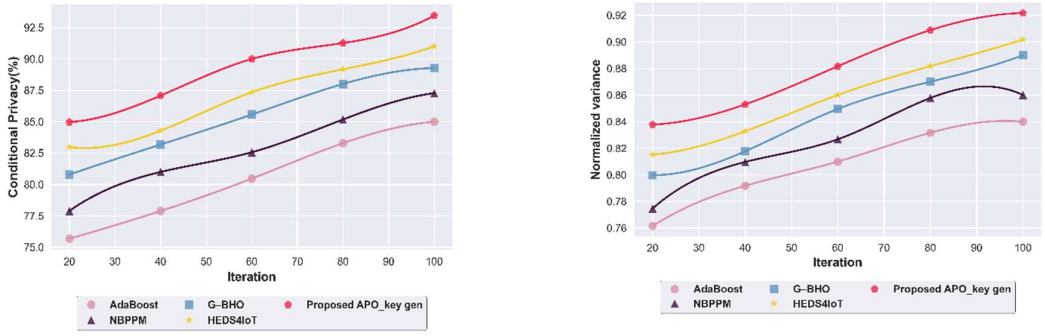
The introduced APO technique is compared with other existing optimal key generation models, like AdaBoost [2], NBPPM [3], G–BHO [4], and HEDS4IoT [5] to demonstrate its efficiency and superiority

5.6 Comparative Analysis

The evaluation is conducted by varying the iteration, to assess the dominance of APO in optimal key generation. The details of this investigation are explained below.

(a) Estimation utilizing the Cleveland dataset

Fig. 5 presents the analysis of APO in optimal key generation based on the Cleveland dataset. Fig. 5 (a) illustrates the validation of APO compared to existing approaches in terms of conditional privacy. The conditional privacy achieved by various methods, including AdaBoost, NBPPM, G–BHO, HEDS4IoT, and APO is 75.678%, 77.896%, 80.788%, 82.976%, and 84.967% respectively for the iteration 30. Notably, the APO approach shows an improvement of 7.071% compared to the NBPPM technique. The validation of APO compared to existing approaches in terms of normalized variance is depicted in Fig. 5 (b). The normalized variance for various methods including AdaBoost, NBPPM, G–BHO, HEDS4IoT, and APO is 0.839, 0.860, 0.889, 0.901, and 0.921 respectively. The results indicate that APO outperforms the AdaBoost method by 8.2%.

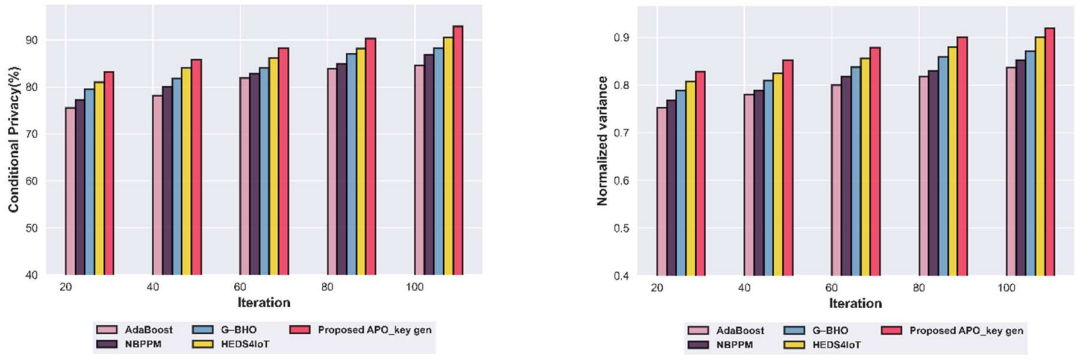


(a) (b)

Fig. 5. *Comparative evaluation of APO utilizing (a) conditional privacy and (b)normalized variance for the Cleveland dataset*

(b) Estimation using the Hungarian dataset

The analysis of APO in optimal key generation based on the Hungarian dataset is signified in Fig. 6. Fig. 6 (a) shows the comparison of APO with existing approaches in terms of conditional privacy. The conditional privacy obtained by various methods including AdaBoost, NBPPM, G–BHO, HEDS4IoT, and APO is 75.574%, 77.277%, 79.567%, 81.004%, and 83.188% respectively for the iteration 30. The APO approach was enhanced by 5.911% compared to the NBPPM technique. Fig. 6 (b) presents the assessment of APO with existing techniques, evaluated based on normalized variance. The newly devised method gained the normalized privacy of 0.919 for iteration 100. The prevailing methods attained the normalized privacy as AdaBoost is 0.836, NBPPM is 0.852, G–BHO is 0.871, and HEDS4IoT is 0.899 for the iteration 100. This demonstrates that the normalized variance is enhanced by 4.8% compared to the prevailing G–BHO method.

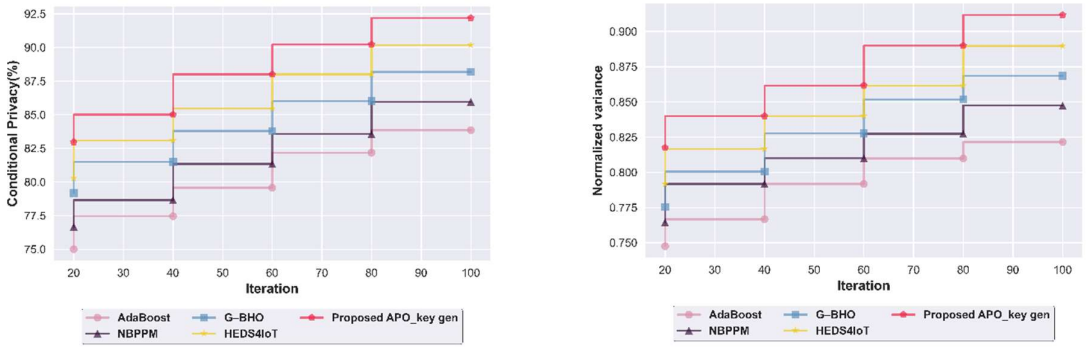


(a) (b)

Fig. 6. *Comparative evaluation of APO utilizing (a) conditional privacy and (b)normalized variance for the Hungarian dataset*

(c) Estimation using the Switzerland dataset

Fig. 7 displays the investigation of APO for optimal key generation based on the Switzerland dataset. The assessment of APO compared to existing approaches in terms of conditional privacy is depicted in Fig. 7 (a). The conditional privacy gained by various methods, including AdaBoost by 75.005%, NBPPM by 76.678%, G–BHO by 79.167%, HEDS4IoT by 80.277%, and APO by 82.969% respectively for the iteration 30. Notably, the APO approach was enhanced by 6.291% compared to the NBPPM technique. Fig. 7 (b) demonstrates the assessment of APO with existing techniques, based on normalized variance. The devised technique achieved the normalized privacy of 0.911 for iteration 100. The prevailing methods attained the normalized privacy as AdaBoost is 0.821, NBPPM is 0.847, G–BHO is 0.868, and HEDS4IoT is 0.889 for iteration 100. This demonstrates that the normalized variance is enhanced by 4.3% compared to the prevailing G–BHO method.



(a) (b)

Fig. 7. *Comparative evaluation of APO utilizing (a) conditional privacy and (b)normalized variance for the Switzerland dataset*

5.7 Comparative Discussion

The results obtained from the APO for optimal key generation for conditional privacy and normalized variance at iteration 100 are shown in Table 1. APO achieved impressive results, with conditional privacy of 93.456% and a normalized variance of 0.921. In comparison, the conditional privacy values for existing methods are AdaBoost, NBPPM, G–BHO, and HEDS4IoT 75.678%, 77.896%, 80.788%, and 82.976% respectively. The normalized privacy for existing methods, such as AdaBoost, NBPPM, G–BHO, and HEDS4IoT, are 0.839, 0.860, 0.889, and 0.901. The outcomes demonstrate that APO excels in providing greater robustness and accuracy The APO method efficiently handles and creates highly unpredictable complex keys making them more resistant to attacks thus ensuring privacy protection.

Table 1. *Comparative discussion*

Datasets	Metrics	AdaBoost	NBPPM	G–BHO	HEDS4IoT	Designed APO
Cleveland	***Conditional Privacy (%)***	84.997	87.277	89.277	91.007	***93.456***
	Normalized Variance	0.839	0.860	0.889	0.901	***0.921***
Hungarian	***Conditional Privacy (%)***	84.997	86.885	88.277	90.477	92.884
	Normalized Variance	0.836	0.852	0.871	0.899	0.919
Switzerland	***Conditional Privacy (%)***	83.857	85.968	88.177	90.177	92.188
	Normalized Variance	0.821	0.847	0.868	0.889	0.911

6. Conclusion

IoT wireless networks and environments facilitate the exchange of information between devices, and these factors can expose confidential and sensitive data to security and privacy breaches. Fortunately, data sanitization mechanisms can help mitigate these risks before sharing and distributing data. However, in large-scale IoT data applications, implementing privacy protection measures can be quite time-consuming, especially when using evolutionary algorithms. To address various privacy challenges and threats in the IoT, this research introduces an effective approach called APO for optimal key generation. This approach aims to enhance the security and privacy of IoT data, making it a valuable contribution to the field. The data owner uploads information to the cloud, converting it to polynomial values for privacy. RV coefficients are identified to assess relationships, and bilinear transformation anonymizes data. An optimal key is generated and encrypted, allowing users to access the original data. The established APO method has demonstrated impressive results, achieving an overall conditional privacy rate of 93.456% and a

normalized variance of 0.921. Future work will focus on integrating a hybrid DL model to enhance the performance and adaptability of newly developed techniques.

Compliance With Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] W. Wang, H. Huang, Z. Yin, T. R. Gadekallu, M. Alazab, and C. Su, "Smart contract token-based privacy-preserving access control system for industrial Internet of Things", Digital Communications and Networks, vol. 9, no. 2, pp.337-346, 2023.
[2] S. El-Gendy, M.S. Elsayed, A. Jurcut, and M. A. Azer, "Privacy preservation using machine learning in the internet of things", Mathematics, vol. 11, no. 16, pp.3477, 2023.
[3] S. K. Jain, and N. Kesswani, "A noise-based privacy preserving model for Internet of Things", Complex & Intelligent Systems, vol. 9, no. 4, pp.3655-3679, 2023.
[4] M. Kumar, P. Mukherjee, S. Verma, Kavita, J. Shafi, M. Wozniak, and M. F. Ijaz, "A smart privacy preserving framework for industrial IoT using hybrid meta-heuristic algorithm", Scientific Reports, vol. 13, no. 1, pp.5372, 2023.
[5] A. Telikani, A. Shahbahrami, J. Shen, G. Gaydadjiev, and J.C.W. Lin, "An edge-aided parallel evolutionary privacy-preserving algorithm for Internet of Things", Internet of Things, vol. 23, pp.100831, 2023.
[6] P. K. Sadhu, V. P. Yanambaka, and A. Abdelgawad, "Internet of things: Security and solutions survey", Sensors, vol. 22, no. 19, pp.7433, 2022.
[7] A. S. Anakath, R. Kannadasan, N. P. Joseph, P. Boominathan, and G. R. Sreekanth, "Insider Attack Detection Using Deep Belief Neural Network in Cloud Computing", Computer Systems Science & Engineering, vol. 41, no. 2, 2022.
[8] R. Ravindra Nikam, and R. Shahapurkar, "Data privacy preservation and security approaches for sensitive data in big data", In Recent Trends in Intensive Computing, pp. 394-408, 2021.
[9] V. Henriques, and R. Malekian, "Mine safety system using wireless sensor network", IEEE access, vol. 4, pp.3511-3521, 2016.
[10] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa, "Insight into insiders and it: A survey of insider threat taxonomies, analysis, modeling, and countermeasures", ACM Computing Surveys (CSUR), vol. 52, no. 2, pp.1-40, 2019.
[11] P. Porambage, M. Ylianttila, C. Schmitt, P. Kumar, A. Gurtov, and A. V. Vasilakos, "The quest for privacy in the Internet of Things", IEEE Cloud Computing, vol. 3, no. 2, pp.36-45, 2016.
[12] I. Wagner, and D. Eckhoff, "Technical privacy metrics: a systematic survey", ACM Computing Surveys (Csur), vol. 51, no. 3, pp.1-38, 2018.
[13] A. George, and A. Sumathi, "Dyadic product and crow lion algorithm-based coefficient generation for privacy protection on cloud", Cluster Computing, vol. 22, pp.1277-1288, 2019.
[14] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, and Q. Arshad, "Data and privacy: Getting consumers to trust products enabled by the Internet of Things", IEEE Consumer Electronics Magazine, vol. 8, no. 2, pp.35-38, 2019.
[15] Heart disease dataset was taken from https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset, and accessed on October 2024.
[16] X. Wang, V. Snášel, S. Mirjalili, J. S. Pan, L. Kong, and H. A. Shehadeh, "Artificial Protozoa Optimizer (APO): A novel bio-inspired metaheuristic algorithm for engineering optimization", Knowledge-Based Systems, vol. 295, pp.111737, 2024.
[17] D. He, S. Zeadally, B. Xu, and X. Huang, "An efficient identity-based conditional privacy-preserving authentication scheme for vehicular ad hoc networks", IEEE Transactions on Information Forensics and Security, vol. 10, no. 12, pp.2681-2691, 2015.
[18] Z. H. A. N. G. Heng-Lei, Y. R. Marangoni, Z. U. O. Ren-Guang, and H. U. Xiang-Yun, "The Improved Anisotropy Normalized Variance for Detecting Non-Vertical Magnetization Anomalies", Chinese Journal of Geophysics, vol. 57, no. 6, pp.800-808, 2014.
[19] D. C. Adams, "Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure", Methods in Ecology and Evolution, vol. 7, no. 5, pp.565-572, 2016.