Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-022-03866-1

ORIGINAL RESEARCH

PPDMIT: a lightweight architecture for privacy-preserving data aggregation in the Internet of Things

Mehdi Gheisari[1] · Amir Javadpour[1,5] · Jiechao Gao[2] · Aaqif Afzaal Abbasi[3] · Quoc-Viet Pham[4] · Yang Liu[1]

Received: 17 September 2019 / Accepted: 26 April 2022

Abstract

Data is generated over time by each device in the Internet of Things (IoT) ecosphere. Recent years have seen a resurgence in interest in the IoT due to its positive impact on society. However, due to the automatic management of IoT devices, the possibility of disclosing sensitive information without user consent is high. A situation in which information should not be unintentionally disclosed to outside parties we do not trust, i.e., privacy-preservation. Additionally, IoT devices should share their data with others to perform data aggregation and provide high-level services. There is a trade-off between the amount of data utility and the amount of disclosure of data. This trade-off has been causing a big challenge in this field. To improve the efficiency of this trade-off rather than current studies, in this study, we propose a Privacy-Preserving Data Aggregation architecture, PPDMIT, that leverages Homomorphic Paillier Encryption (HPE), K-means, a One-way hash chain, and the Chinese Remainder Theorem (CRT). We have found that the proposed privacy-preserving architecture achieves more efficient data aggregation than current studies and improves privacy preservation by utilizing extensive simulations. Moreover, we found that our proposed architecture is highly applicable to IoT environments while preventing unauthorized data disclosure. Specifically, our solution depicted an 8.096% improvement over LPDA and 6.508% over PPIOT.

Keywords Data aggregation · Internet of Things · Privacy preservation · Paillier encryption · Low-cost

1 Introduction

Nowadays, the Internet of Things (IoT) has become increasingly popular in many aspects of our lives. Industry and government are taking steps to address the risks associated with IoT technologies and products, especially in the areas of privacy and security, and guidelines, standards, and regulations have been developed regionally and internationally to deal with these concerns (Gheisari et al. 2021; Gheisari et al. 2020). IoT evolved after passing some technologies such as RFID, embedded systems, and wireless sensor networks that aim to sense the environment effectively to make high-level decisions. Wireless networks under (software-defined networking) SDN management and the use of Internet of Things networks are very popular today. In addition, companies and wireless stations use them for specific purposes, and security issues and checking their privacy are very important. Each IoT device should produce data over time and collaborate with others (Javadpour et al. 2020a, 2018; Mirmohseni et al. 2020; Javadpour and Wang 2021; Javadpour 2019a, b).

✉ Amir Javadpour
a.javadpour87@gmail.com; a_javadpour@e.gzhu.edu.cn; amirjavadpour@cs.hitsz.edu.cn

✉ Jiechao Gao
jg5ycn@virginia.edu

Mehdi Gheisari
mehdi.gheisari61@gmail.com

[1] Department of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China

[2] Department of Computer Science, University of Virginia, Virginia, USA

[3] Department of Software Engineering, Foundation University, Islamabad 44000, Pakistan

[4] Korean Southeast Center for the 4Th Industrial Revolution Leader Education, Pusan National University, Busan 46241, Korea

[5] ADiT-Lab, Electrotechnics and Telecommunications Department, Instituto Politécnico de Viana do Castelo, 4900-347 Porto, Portugal

Published online: 17 June 2022

Sensitive information about a building, such as the number of people alive there, may be compromised by this collaboration. To avoid possible harm in the future, sensitive data must never be unintentionally disclosed. On the other hand, high-level services such as the aggregation of data must be shared. Since a large number of IoT devices are performing their actions autonomously, automatic privacy-preserving solutions are of importance for IoT systems. Proposed privacy-preserving solutions should be lightweight because many IoT devices are resource-constrained (Gheisari et al. 2018; Badra and Zeadally 2017).

On the other hand, "Cloud Computing" (CC) can support IoT environments because it can provide unlimited computing and storage and great virtualization of IoT devices. Cloud computing can be integrated with the IoT to achieve more effective environments, i.e., IoT-Cloud. It is worth mentioning that research on IoT Cloud is still in its infancy stage (Javadpour et al. 2018).

This paper proposes a lightweight solution for IoT Cloud environments to preserve the privacy of sensed sensitive data more efficiently and achieve efficient data aggregation. This is achieved through leveraging several methods such as homomorphic Paillier encryption for providing secure data transfer connection while manipulating encrypted data is possible, a one-way hash chain with the aim of early false data detection, Gaussian distribution algorithm to find valuable data among all collected data while it is lightweight, and Chinese remainder theorem for data aggregation (Guan et al. 2019; Javadpour et al. 2020b). In our proposed solution, Privacy-Preserving Data Management in the Internet of Things (PPDMIT), IoT devices send their data to the Cloud via homomorphic paillier encryption, after which the remote Cloud removes false information and finds valuable information, aggregating it. Data aggregation involves condensing data for the purposes of statistical analysis and then expressing it on a condensed basis. Several sources of data are combined into one cluster using data aggregation tools. Our data aggregation methods can give us new insights and help us discover new relationships. The process involves several input packets being received by intermediate nodes such as gateways. Following aggregation, the network will produce one output packet. Based on IoT devices, data aggregation is often used to gain more insight into particular groups of people.

This paper proposes a lightweight privacy-preserving method for IoT Cloud environments, namely, PPDMIT. Our proposed privacy-preserving architecture is put to the test by running simulations in an IoT environment.

The key contributions are:

- To achieve PPDMIT, at first, each device uses a homomorphic paillier encryption method to send its data to the remote Cloud. And one-way hash chain method is used in the CC for early Detection and removing false data.
- We adopt a Gaussian distribution algorithm, K-means, in the Cloud for finding valuable data.
- We use Chinese Remainder Theorem for data aggregation over hybrid IoT devices.

In this way, the redundancy in a network can be reduced, protecting privacy. The remaining section of this paper is as follows. We describe relevant background info on the Internet of Things privacy-preserving methods in regard to IoT in Sect. 2. Section 4 presents the components of the solution. Section V describes the proposed PPDMIT architecture, which is evaluated in Section VI. Section VII summarizes the paper and describes future work.

2 Background

In this section, we ascertain the necessary IoT, information privacy-preserving in IoT, and common privacy-preserving methods.

2.1 Internet of Things

One major component of IoT is wireless sensor networks that are connected through the Internet; some differences with WSNs are:

- IoT covers all types of objects such as humans, handwriting, PCs, and so on.
- In pure IoT, routing is not implemented because each device sends its data directly to the Internet without any broker.
- A whole wireless sensor network can be considered as one node in IoT.

IoT is not a single technology but a set of technologies that are participating in local activities and interacting with each other. The Internet is used to connect all aspects of our lives worldwide to provide high-quality services. IoT affects how we live and how we work and how we can achieve better performance. Whenever and wherever devices are connected, a high level of security should be assured. In addition to physical items, these devices can include vehicles, home appliances, etc. These devices should be equipped with sensors actuators. It is anticipated that billions of devices will exist by 2025 (Gheisari et al. 2021).

Figure 1 shows the Internet of Things concept from a schematic point of view. As Fig. 1 shows, all devices such as smart cameras are connected to provide more humanized services in the IoT era. As mentioned above, a large number

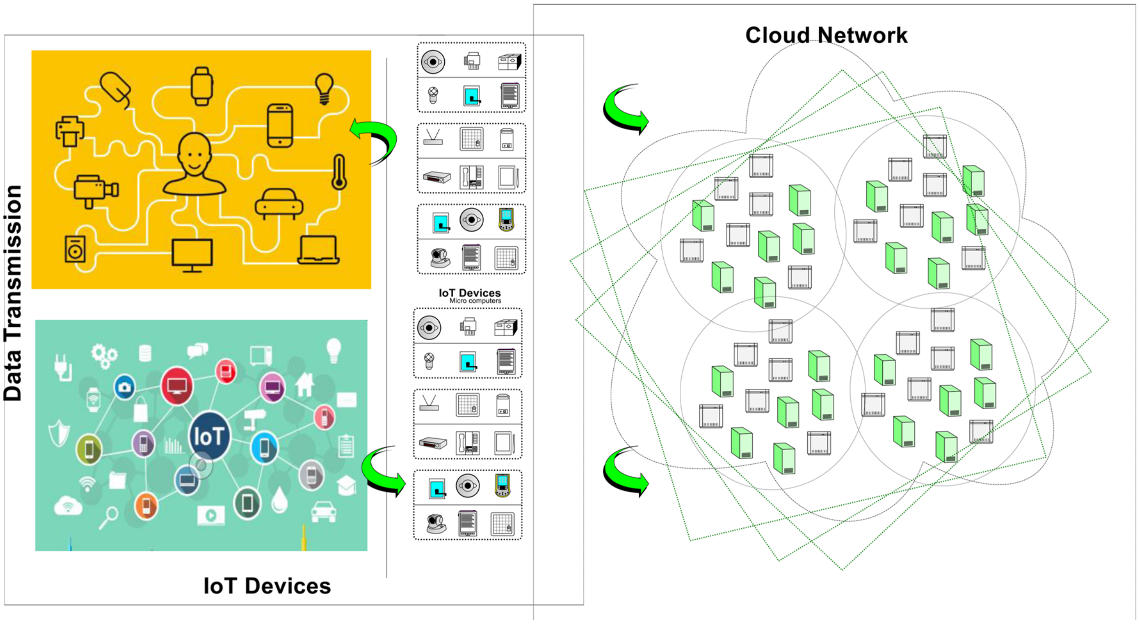


Fig. 1 The details of Internet of Things and Cloud network-based (Gheisari et al. 2021; Javadpour and Wang 2021)

of connected devices will exist such that each one produces data over time, thus leading to Big Data.

2.2 Privacy preservation in IoT

Over time, IoT devices produce data increasingly. Data should be managed efficiently to provide a better quality of service (QoS). In other words, IoT devices should share their data and collaborate with others. One technology that can be applied to obtain better QoS is data aggregation. We can discover patterns and convert raw data into useful knowledge based on data aggregation. However, sensitive data may be disclosed. Thus, we should consider solutions that prevent unintentional disclosure of data to adversaries to reduce the possibility of misusing data and harming the system (Rachels 2017).

Sensitive information can be divided into three main subcategories:

- Personal: e.g., social security number (SSN).
- Sensitive: e.g., Salary and disease.

For example, zip code and age are quasi-identifiers. What is the importance of quasi-identifiers? As a result, quasi-identifiers must be kept private since we can identify individuals by joining data obtained from various external sources, including public voter registration information, hospitals, and news. To prevent the misuse of sensitive data, these three types of data should remain private. The short and simple answer is that we must protect sensitive data from parties we do not trust and do not want access to. In addition, we can distinguish between two types of privacy-preserving concepts: (1) protection of the content; and (2) protection of the context. (Liu et al. 2016). Content protection means protecting sensitive generated data from unintentional disclosure, i.e., we have to provide users a way to process the data so that no one can find the original generated sensitive data when we are facing adversaries. In reverse context, data refers to protecting the information of non-sensed data from information leakages such as sensing time and sensing location.

In more detail, privacy types can be divided into four main subcategories (Javadpour 2019a, b; Gheisari et al. 2018):

- Data privacy: preserving the privacy of produced data
- Location privacy: keeping the locations of IoT devices private.
- Time privacy: keeping the sensing time private.
- User privacy: maintaining user behavior private.

Based on the IoT Security Threat Map that was published in 2017 by the Beecham research group (Badra and Zeadally 2017; Khan et al. 2019), security challenges, as shown in Fig. 2, are becoming worse over time, and we need to address them. The figure shows that damages that may stem from attackers mainly occur in three areas: attacks on platforms, networks, and edge devices. Recently, most attacks tend to IoT Inter-Sector interactions part of IoT

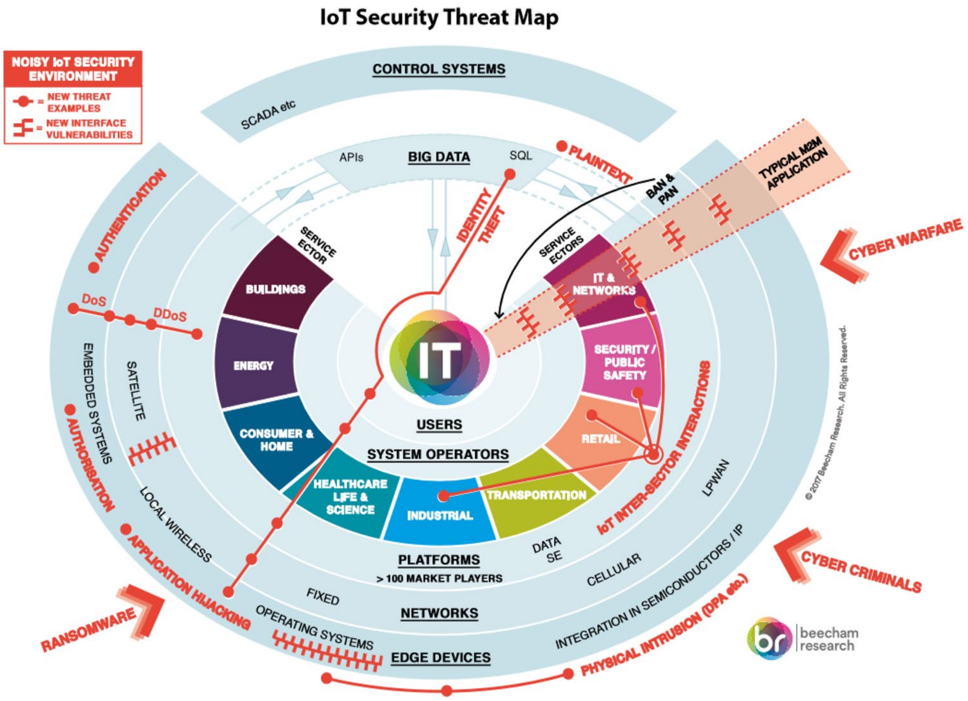


Fig. 2 IoT Security Threat Map (Beecham 2021)

Table 1 Privacy-preserving techniques (Aldeen et al. 2015)

Privacy-preserving techniques	Attributes
Data distortion	Contains data operations like perturbation, blocking, aggregation, merging, swapping, sampling
Data or rules hidden	Hiding sensitive data or rules
K-anonymity	Processing the anonymization procedure of data
L-diverse	Maintains diversity of sensitive attributes by keeping at least K group sizes
Taxonomy tree	Limiting information leakage with the help of tree
Randomization	For example, adding random noises to data based on a logic

environment such as buildings' data, health care data, and so on (Ding et al. 2013).

2.3 Privacy-preserving methods

These methods are described in this section as ways of preserving privacy. Six privacy-preserving solutions are available in environments that deal with large amounts of data and want to maintain privacy. An explanation of the method is provided in Table 1.

3 Related works

In this section, we describe the literature that tries to perform aggregation while keeping the privacy of data.

Martonosi in Martonosi (2016) explained in detail the concerns of how to store data in IoT and cloud environments in terms of security and privacy while they are prepared to be aggregated. They explained the policies followed by the U.S. government for handling privacy preserved. Various technologies and tools for different kinds of applications are discussed. However, they did not show any data aggregation methods for IoT devices while preserving data privacy.

In Zhu et al. (2014), based on each participant's secret key, the authors proposed an efficient algorithm to aggregate data with privacy preservation. There is, however, one disadvantage, namely, the inexcusable degree of privacy preservation. In other words, if the secret key is penetrated one time, it would be easier for the rest of the attacks to misuse the system's vulnerability. Moreover, their solution is not able to remove false data and is not effective in aggregating hybrid IoT devices' data.

To achieve better data aggregation while preserving privacy, authors in Ruj and Nayak (2013) have been employed homographic paillier encryption. One of its main restrictions is that the method only depends on encryption to provide privacy from outsiders that cause unacceptable privacy-preserving levels. Moreover, their solution can address the diversity among devices effectively.

Raju et al. (2009) applied the homomorphic encryption method to multiplying protocols to preserve privacy. However, it has some drawbacks, such as an unacceptable privacy-preserving degree. Moreover, their solution is not robust to false data generated by adversaries.

As described in the author's presentation Zhang et al.. (2020), Edge computing provides efficient computing and data storage in IoT systems. This data-driven architecture is designed to prevent privacy leakage from other incompatible entities by protecting the privacy of user-side data. For a wide range of IoT applications utilizing cloud computing, they propose Privacy Data Collection (PPDA) schemes. However, PPDA solutions are not suitable for edge computing because IoT smart devices have high performance and privacy requirements. LVPDA, a lightweight, verifiable PPDA design, addresses this challenge by combining Paillier homomorphic encryption with an online/offline signature method to ensure privacy as well as comprehensive integration verification. Compared with other PPDA methods, our design implements a lightweight PPDA with lower computation complexity.

IoT-based smart grids could collect and transmit data with more accuracy and frequency than traditional networks, according to a study by Wang et al.. By enabling efficient data source authentication, ensuring the integration of users, and ensuring dynamic exit PDAM, a privacy data aggregation scheme for IoT-enabled smart grids aims to resolve these issues. As long as PDAM is implemented, users can be totally protected against external and internal attackers, malicious collectors, curious controllers, and malicious collectors. They have developed PDAM to meet many known security requirements and to provide optimal performance for a smart grid system. Additionally, experiments and comparative studies have shown PDAM to be superior to other recent proposals (Wang et al. 2021).

Depending on data quality, the Internet of Things will be secure and tolerant of errors. The author Q. Wang et al. (2021) shows how this can be accomplished. There is a multitude of existing schemes whose data collection is not filtered. In addition to privacy concerns, aggregation schemes pose numerous other challenges, such as lightweight requirements and tolerable requirements. A number of numerical and Boolean answers to queries can be provided with PLSA-FT, depending on the conditions of the cloud center. Additionally, Paillier Hemorrhoid Initial Encryption ensures data privacy and provides fault tolerance for IoT malfunctions. Based on their conclusions, PLSA-FT protects confidential information, ensures privacy, authenticates sources, verifies integrity, ensures fault tolerance, and manages dynamic membership.

Mukkamala et al. (2011) compared a Fuzzy-based approach of mapping. The authors combined several values into one singular value for a more efficient process. The combination brings privacy while saving the bandwidth of the network. One of its drawbacks is its high computational cost.

In Kamakshi and Babu (2012), the authors proposed a novel idea for identifying sensitive attributes automatically, and then the data is modified so that the original properties of the data are preserved. One of its notable drawbacks is that their method cannot be generalized to cover a variety of domains. Moreover, the authors did not calculate the amount of overhead of their solution.

Moreover, authors in Lai et al. (2014) proposed an outsourcing association rule mining that is approximately secure and preserves privacy by leveraging both data privacy and mining privacy. One advantage is that the solution enables false data to be identified in the mining process. However, one disadvantage is that it is non-deterministic to adversaries in cloud servers that sometimes causes high computational cost.

In (Lu et al. 2017), Lightweight Privacy Data Aggregation (LPDA) is an approach to tackle this challenge proposed by the authors for fog computing-enhanced IoT. As well as aggregating data from hybrid IoT devices, it also filters early injected false data by using homomorphic Paillier encryption, the Chinese Remainder Theorem, and a one-way hash chain technique. By using differential privacy techniques and comprehensive evaluations for security and privacy enhancement, LPDA has proven to be both secure and private.

Authors in Lai et al. (2014) described a homomorphic public key encryption scheme for binary digits. They developed a PIR protocol that reduces the data strikingly to achieve data privacy (Melchor and Gaborit 2008). The authors used an election system to check the validity of ballots given by users. They suggested a 2DNF protocol which is described for safety from malicious users and preserving the privacy of users. One of the most striking drawbacks is that this work did not consider false data and/or save the network's bandwidth.

Tassa et al. (2013) described a distributed protocol based on association rules in Databases distributed on multiple servers that are spread horizontally. One of its drawbacks is that they did not propose an effective protocol for disparity verification, and they also did not consider the amount of communication cost. Fortunately, their system is approximately accurate.

In (Aïvodji et al. 2019), authors have presented several additional challenges in terms of privacy and security. They have proposed a new architecture for smart homes, the IOT-FLA, combining Federated Learning and secure data aggregation while focusing on security and privacy. In achieving more security and privacy in smart homes, we hope that our proposal will be a step forward.

Zhang et al.. (2013) have been proposed a method for anonymity; one privacy-preserving method depends on an efficient quasi-identifier index. They also tried to protect privacy when new data is added to the data set. One of its drawbacks is that this work did not consider false data. Moreover, they did not check whether or not the data was valuable.

To fill the gap of current studies, we propose PPDMIT to provide efficient IoT environments that preserve the privacy of generated data while performing data aggregation over hybrid IoT devices' data. Besides, PPDMIT considers both false and valuable data to obtain efficient and clean IoT environments.

Here, we investigate the literature more deeply and compare it with (Networks Using Pascal encryption, considering data privacy, Using Data Management, Clustering for finding valuable, cost, and complexity in Table 2.

Table 2 Privacy-preserving techniques (Aldeen et al. 2015)

References	Network (big data hybrid IoT)	Paillier encryption	preserving data privacy	Data management	Clustering for finding valuable	lightweight	Cost an d complexity
Martonosi (2016)	Y	–	Y	Y	N	N	High
Zhu et al. (2014)	N	–	Y	Y	N	N	Low
Ruj and Nayak (2013)	N	–	–	N	N	Y	High
Raju et al. (2009)	N	-	-	Y	N	Y	Low
Zhang et al. (2020)	Y	Y	Y	Y	N	Y	Low
Wang et al. (2021)	Y	Y	Y	Y		N	Low
Wang and Mu (2021)	Y	Y	Y	Y	Y	Y	High
Mukkamala and Ashok (2011)	N	–	Y	N	N	N	Low
Kamakshi and Babu (2012)	N	–	–	N	N	N	High
Lai et al. (2014)	N	–	Y	Y	Y	N	Low
Lu et al. (2017)	Y	Y		Y	N	Y	High
Melchor and Gaborit 2(008)	Y	–	–	Y	–	N	High
Tassa (2013)	N	–	–	Y	No (association rules)	Y	Low
Aïvodji et al. (2019)	N	-	-	N	Y	N	High
Zhang et al. (2013)	Y	-	Y	Y	N	N	High

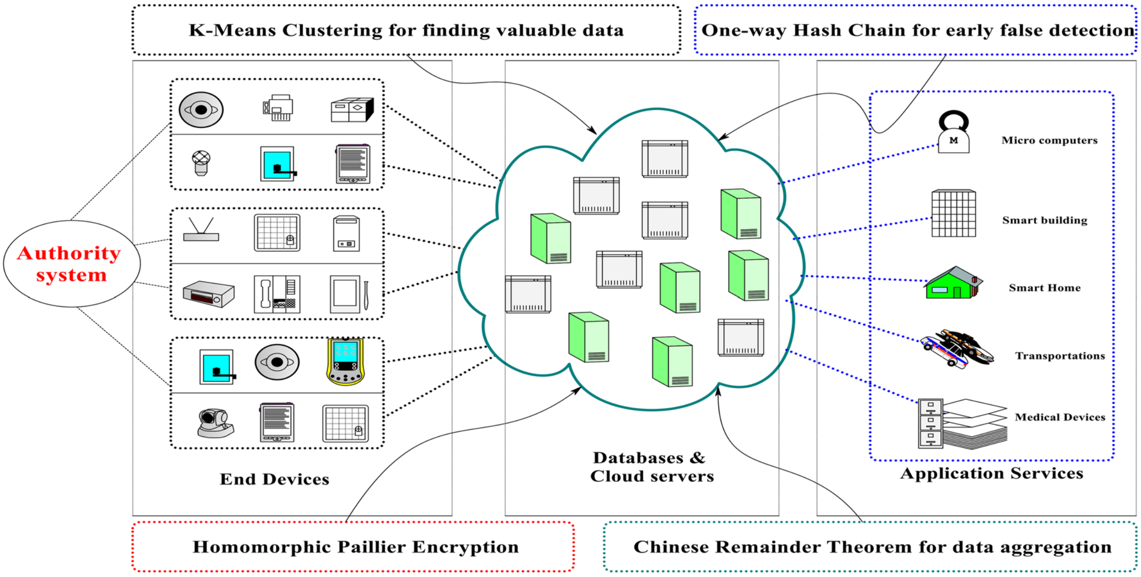

K-Means Clustering for finding valuable data
One-way Hash Chain for early false detection
Authority system
M
Micro computers
Smart building
Smart Home
Transportations
Medical Devices
End Devices
Databases & Cloud servers
Application Services
Homomorphic Paillier Encryption
Chinese Remainder Theorem for data aggregation

Fig. 3 Describing the process of Privacy-Preserving Data Management in IoT

4 Fundamental components of ppdmit

This section pays attention to the necessary background information for designing our solution. Figure 3 examines the Privacy-Preserving Data Management process in the IoT. Which includes different sections such as K-Means Clustering for finding valuable data, One-way Hash Chain for early false detection, Chinese Remainder Theorem for data aggregation, and Homomorphic Paillier Encryption.

4.1 Homomorphic cryptosystem

It makes it possible to perform complex calculations on encrypted data without compromising the encryption with homomorphic cryptosystems. This cryptosystem is based on the decisional composite residual assumption. Simply and straightforwardly, we can perform operations on encrypted data and get the encrypted result that, when decrypted, would be the same as we would get if the operations were performed on the decrypted text in the first place. Keeping data secret from others is the main purpose of data transformation. Data security can also be ensured with encryption. A sensible approach would be to see homomorphic encryption as transparent and secure so that manipulators can manipulate the data without stealing it. In addition, the data is transparent so that any manipulator can look at the data without being able to manipulate it. In brief, the third party can manipulate the received encrypted data and send it back to the data owner while manipulating it without allowing anyone else to understand it (Zheng and Huang 2013). The encrypted information of m_1 and m_2 can be computed from the public key when the $m_1 + m_2$ encryption key is missing. Below is a detailed explanation of how it works.

- It will be necessary to simultaneously and independently pick two large prime numbers p and q so that $(pq, (p-1)(q-1)) = 1$. These prime numbers must also be equal in length.
- It will be determined by calculating $n = pq$, and, $\lambda = 1\ cm(p-1, q-1)$ from 1 by using Least Common Multiple.
- The random integer is selected by in Z
- By checking the existence of the modular multiplicative inverse, we ensure that n divides the order of g: $\mu = (L(g^{\lambda} \mod n^2))^{-1} \mod n$
- L is the definition of a function $L(x) = \frac{x-1}{n}$
- A public key consisting of (n, g) is required for encryption and a private key consisting of (λ, μ) is required for decryption.

By exploiting certain discrete logarithms that are easily computed, Paylier's cryptosystem works. By using the binomial theorem as an example,

$$(1+n)^x = \sum_{k=0}^{x} \binom{x}{k} n^k = 1 + nx + \binom{x}{2} n^2$$

As a consequence of the theorem, we have: $(1+n)^x \equiv 1 + nx \pmod{n^2}$ and $y = (1+n)^x \mod n^2$, then: $x \equiv \frac{y-1}{n} \pmod{n^2}$.

Accordingly:

$$L((1+n)^x \mod n^2) \equiv x \pmod{n},$$

The definition of the function L is the (Quotient of Interinteger Division).

Homomorphic properties make the system malleable, however. This security feature does not protect against adaptive chosen DDos and does not offer the highest level of semantic security. Cryptographic strength is usually not malleability. Although an IoT system and cryptographic threshold protocol might not require this property, other applications might.

4.2 One-way hashing

This is the most famous cryptographic primitive as it involves successively building up one-way hashes of data. One-way functions are algorithms that map data of arbitrary size to strings with a fixed size, and it is impossible to find and invert the original data given the algorithm's formula. Hashing is a form of data protection that is used to secure strings and files. By hashing, the information can be converted to a digest message or a hash, which is a number derived from a string or text. These available digests make it easy to verify whether the sent and received messages have the same hash and no tampering. Call numbers booked using libraries within a domain are an example of a hash function. A library's books are identified by their unique call numbers so that it is possible to locate them by their call number. A hash function returns a unique hash number that is called a universal hash function to verify data. In hashing, the received data and digests cannot be reversed to the original one, for example, MD5 in Mendel et al. (2009). The one-way hash function can also be applied in various domains such as database indexing, caching, program compilation, error checking, or false data selection. The function is one-way, meaning that it cannot be inverted or reversed. When searching for messages that produce a given hash, most people use brute-force searching to see if different inputs produce a match or use rainbow tables of matched hashes. Modern cryptography relies on cryptographic hash functions. A hash function's advantages are that it enables insertion, deletion,

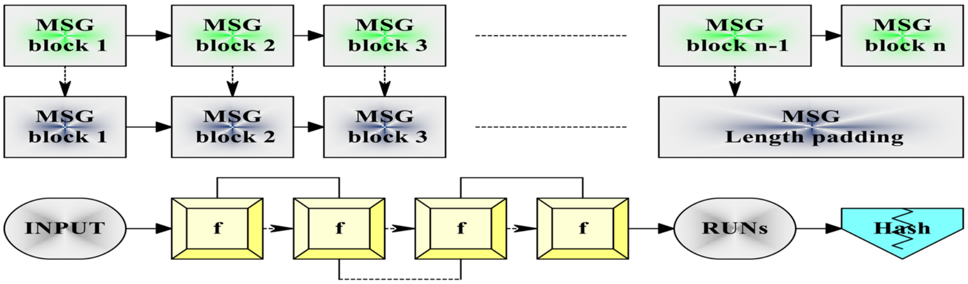


Fig. 4 Optimal cryptographic hashing requires deterministic hashing

and retrieval simultaneously. IMN devices, which are able to filter false data, can assist in fighting PPDMIT false data injection attacks(Fig. 4) (Jho et al. 2005).

4.3 Chinese remainder theorem

The Chinese remainder theorem is a theorem of number theory, which discovers the remainders of the Euclidean division of an integer N by several integers, then one can determine the remainder of the division of a number N uniquely by the product of these integers, provided that the divisors are pairwise coprime (Erdos and Schönheim 1969). An important calculation algorithm in modular arithmetic which enables a person to solve simultaneous equations concerning different moduli in considerable generality. In brief, it addresses such problems of finding a number that, for example, leaves a remainder of 0 when the number is divided by 5, the remainder 6 when the number is divided by 7, and the remainder 10 when it is divided by 12. The simplest result is 370. It is notable to mention that the result is not unique since any multiple of $5 \times 7 \times 12 = 420$ can be added to it, and the result will still satisfy the problem.

Furthermore, it is widely acceptable for large computing on large devices because it allows replacing the limitation on the size of the result with several similar operations on small integers. It provides a unique solution to simultaneous linear congruence. We use it for data aggregation over massive IoT devices in our scenario. Chinese remainder theorem states that, given a n, a series of remainders may be obtained by division by several integers based on Euclidean division. Under the condition that the divisors are coprimes, it is possible to determine the remainder of n by computing the product of these integers. In this way, as long as it is clear a bound on the size of the result, a computation can be replaced by a few smaller computations. We may call these integers n_1, ..., n_k moduli or divisors because they are greater than 1. We may call these integers moduli or divisors because they are greater than 1. Because these integers are greater than 1, we can call them moduli or divisors. Let us denote the product of n_i by N.

For example, if a_1, ..., a_k are any entire integers, and n_i are pairwise coprime, then the system is congruent:

$$x = a_1 \quad (\quad \mathrm{mod}\ n_1)$$
$$\dots$$
$$x = a_k \quad (\quad \mathrm{mod}\ n_k)$$

The congruence between two solutions to a problem is modulo N.

$$x_1 = x_2(\quad \mathrm{mod} \quad N \quad for(all))$$

The remainder of the Euclidean division of x by each n_i can be used to determine whether a value x is a solution. It is sufficient to check each integer between 0 and N until the solution successively is found for this problem. Despite being straightforward, this method is extremely inefficient. The solution must be checked for integers (including 0). As an example considered here, we check t integers (including 0). There is a constant factor in the size of the input, so the input is up to N digits, and there are an average of N operations in the calculation. We are therefore trying to find a polynomial $P(X)$ that is congruent with:

$$P(X) \equiv A_i(X).\left(\mathrm{mod} P_i(X)\right) = \prod_{i=1}^{k} P_i(X).Q_i(X)$$
$$= \frac{Q(X)}{P_i(X)} \quad for \rightleftarrows i = 1, \dots, k$$

When $\frac{1}{Q(X)}$ is decomposed into partial fractions, we obtain $S_i(X)$ polynomials of degree $\bigcap_{i=1}^{n} X_i S_i(X) < \bigcap_{i=1}^{n} X_i d_i$ with degrees F, as shown below.

$$\frac{1}{Q(X)} = \sum_{i=1}^{k} \frac{S_i(X)}{P_i(X)} \quad \rightarrow 1 = \sum_{i=1}^{k} S_i(X)Q_i(X).$$

Therefore, by dividing by the polynomial, one gets the solution to the simultaneous congruence system:

$$\sum_{i=1}^{k} A_i(X)S_i(X)Q_i(X) \qquad 1 \le i \le k.$$
$$\sum_{i=1}^{k} A_i(X)S_i(X)Q_i(X) = A_i(X) + \sum_{j=1}^{k} (A_j(X) - A_i(X))S_j(X)Q_j(X) =$$
$$A_i(X).\left(\mathrm{mod} P_i(X)\right),$$

The degree of a solution with degree greater than $D = \sum_{i=1}^{k} d_i$ may be greater than one. The unique solution with less than degree D is obtained by dividing $B_i(X)$ by $A_i(X)S_i(X)$ by $P_i(X)$ and finding the remainder. Hence, we have the solution.

$$P(X)_{all} = \sum_{i=1}^{k} B_i(X)Q_i(X)_{all}$$

5 The proposed method—PPDMIT

This section aims to describe the PPDMIT, a model for IoT environments for data aggregation while preserving the privacy of sensitive data.

5.1 Algorithm of PPDMIT

The algorithm of the PPDMDIT in schematic form is depicted in Fig. 5. The figure shows our proposed method for performing data aggregation while preserving privacy in the Cloud-based Internet of Things environment from a schematic point of view. Intending to achieve data aggregation while preserving data privacy efficiently, PPDMIT leverages four steps. It is notable to mention that the last three steps (i.e., steps 2, 3, and 4) are deployed in the cloud environment, while step 1 is applied at the IoT device level. In more detail, in step 1, IoT devices hash their data, apply the homomorphic paillier encryption method to their data, and send them to the Cloud. In Step 2, before data aggregation is performed by leveraging the Chinese Remainder Theorem, some refinements should be done; for example, the cloud server applies a one-way chain method to remove false data. In step 3, the cloud server applies K-means clustering algorithm to differentiate valuable data from non-valuable ones. Finally, in step 4, the cloud server does aggregation using the Chinese Remainder Theorem (Fig. 5 or Algorithm 1).

6 Performance evaluation

In this section, we simulate PPDMIT to verify its performance in OpenIoT. Open IoT bridges this gap between semantics and data computing, so IoT applications in the

Algorithm 1: PPDMIT

```
The data is contained in the window: Identification and Query for IoT devices.
            The query result has been encrypted using users' private key
                        Leverages is provided to the data access layer by the user
One-Way Hashing: An algorithm that maps an arbitrary size of data to a cryptographic hash function (CHF)
            Homomorphic Cryptosystem: Run Gaussian distribution algorithm in the cloud for finding valuable data.
               K-means to find valuable data
                    ➢   Choosing K will determine the number of clusters.
                    ➢            Next, choose random points from K. These may be different from the input data.
                    ➢                        The center of each point should be determined. The centroid will be used to form K clusters.
                    ➢            To determine the centroid, simply calculate the variance.
                    ➢            In the third step, assign each point to its nearest centroid.
                     Run: homomorphic paillier encryption for encrypting IoT devices
                          The data has been entered: Identification, Query
                                Results of an Q(*) query encrypted using users' PINs
                                      Calculations will be performed by the data access layer
            Chinese Remainder Theorem: each device uses homomorphic paillier encryption
                             ▪   $N$, by several integers, $n_1, \ldots, n_k$ moduli or divisors , denote the product of $n_i$ by $N$ .
                             ➢   $P(X) \equiv A_i(X).(\bmod P_i(X)) = \prod_{i=1}^{k} P_i(X).Q_i(X) = \frac{Q(X)}{P_i(X)} \quad for \quad \ldots \quad i = 1,\ldots,k$
                                          ➢   $\frac{1}{Q(X)} = \sum_{i=1}^{k} \frac{S_i(X)}{P_i(X)} \quad \rightarrow \quad 1 = \sum_{i=1}^{k} S_i(X)Q_i(X).$
                                                   ➢   $P(X)_{all} = \sum_{i=1}^{k} B_i(X)Q_i(X)_{all}$
                  for sending data to the remote cloud
                              one-way hash chain method is used
                                          detection and removing false data.
                                          one-way chain for early false detection
                  Server computes : The data access layer will pass to n servers
                              The layer that computes data access: according to the query, x data points
            will be queried
            The data access layer will perform calculations
                 Computed costs of preserving privacy: CPU usage
```

Fig. 5 Algorithm PPDMIT leverages for privacy-preserving data management in IoT

Cloud can have unified semantics. Open IoT is a generalized standard model for semantic unification of diverse IoT systems by using the W3C Semantic Sensor Network ontology (SSN). Using its infrastructure, all I/O devices are able to gather and annotate data in a semantic manner. Depending on how similar the data sets are, Open IoT can also provide a special feature that permits easy linking. In this manner, it is capable of dealing with data streams compatible with mobile sensors without needing extraneous interfaces. In addition to providing a wide range of tools, it supports cloud-based IoT applications, hence reducing the programming effort. Global Sensor Networks (GSN) is an open-source project on Github that provides open-source libraries for Open IoT. As one of the top ten open source projects, Open IoT has been honored with the Black Duck Award. We did simulation using OpenIoT software (Jayaraman et al. 2017) (Fig. 6). In the simulation setup, 100 homogeneous temperature devices are producing sensitive data while sending their data to the Cloud for data aggregation. Each device senses a random value between 10 and 100. IoT devices are assumed to have unique identification numbers that are natural integer numbers less than 101. In the first step, each device calculates the remainder of its value by its unique ID. Then, they hash their data to the data range between 110 and 220. The data range numbers are based on our assumption, and we note that other ranges of numbers can be used instead. Therefore, more investigation is needed to find the best values. After hashing the values, IoT devices apply Microsoft SEAL (homomorphic encryption scheme), an open-source library that provides a set of encryption libraries to perform computations directly on the encrypted data. A note of significance is that we used the Brakerski/Fan Vercauteren (BFV) homomorphic encryption scheme as an assumption, and additional research is required to determine the best option.

After enabling the manipulation of encrypted data by applying the homomorphic encryption method, IoT devices send their data to the cloud server. The cloud server firstly applies the one-way hash function to find false data. Therefore, we can resist unwanted adversaries that are trying to inject false data into the system. In the next step, step 3, the cloud server applies the K-means classification algorithm to find valuable data from non-valuable ones with $k = 3$. In other words, we have two classes: valuable data and non-valuable data that are determined by K- means algorithm. In our scenario, data is more valuable if it is used in more than 80 percent of the occurrences. Shortly and straightforwardly, if the usage number of IoT device data is higher than 80 percent, it is considered valuable data. It is notable to mention that since our scenario does not involve many IoT devices, we use K-means. In the case of a large volume of data, we need to use a stronger classifier, and we leave it as future research work. Sixty-six out of one hundred IoT devices' data is valuable that can be used for data aggregation as step 4 with leveraging Chinese remainder theorem (CRT). As a result, our solution can aggregate encrypted IoT device data while preventing false data injection and removing non-valuable data. For evaluating PPDMIT, we calculate two important evaluation metrics that are: (1) the computational cost to find whether or not devices can afford it, and (2) the number of unwanted disclosures of data.

Figure 7 shows the amount of the computational cost of the system. It is evident that the amount of PPDMIT overload in average is around 35 percent. Although 35 percent is affordable for many IoT devices, it would be better to use

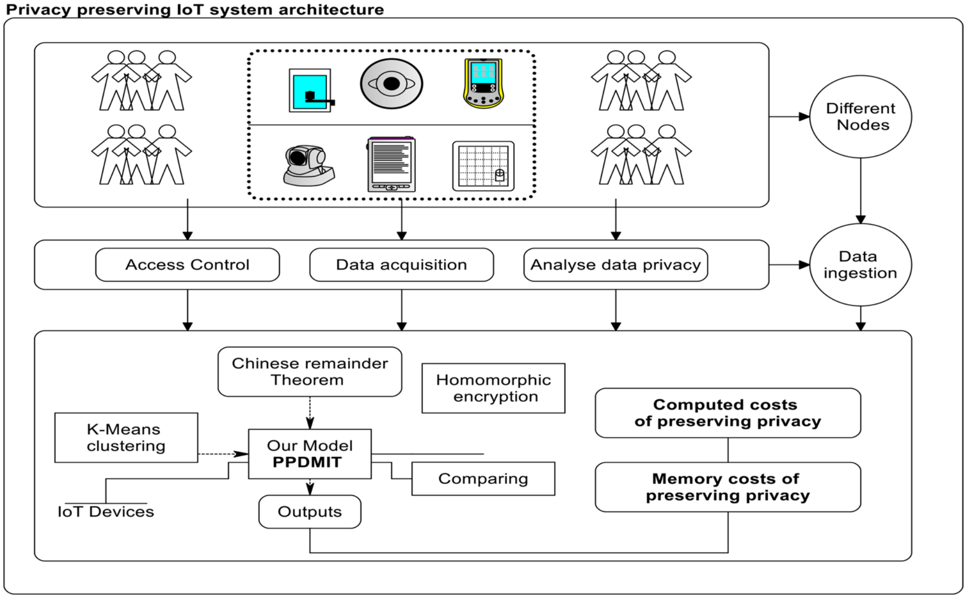


Fig. 6 Details of implementation and simulation in OpenIoT

Fig. 7 The details of computed costs of preserving privacy, extra CPU usage average in different nodes

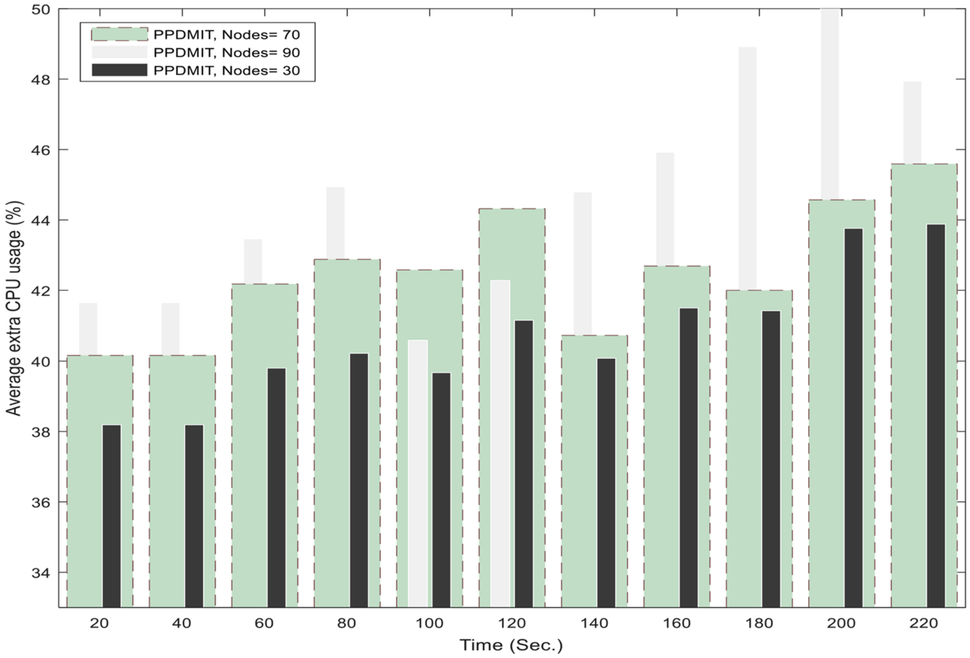
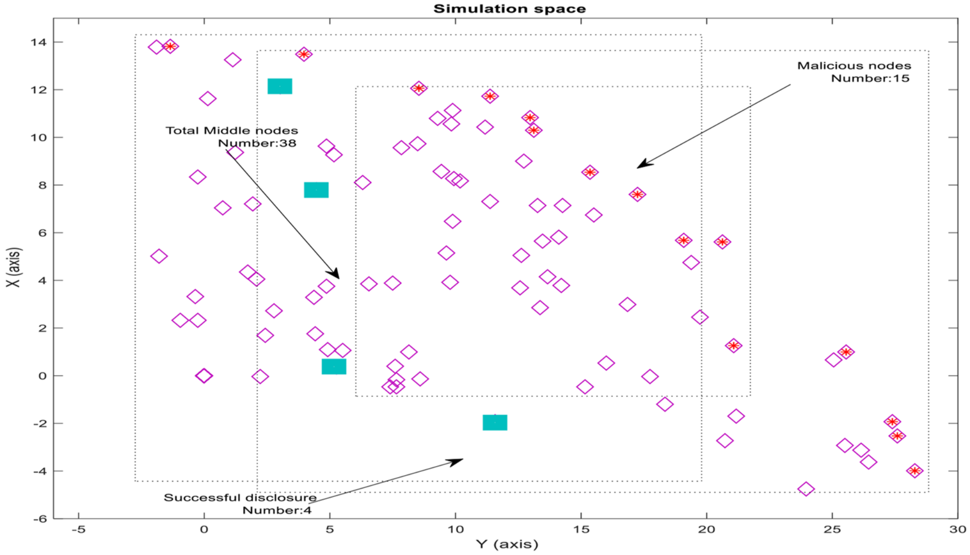


Fig. 8 Number of unintentional disclosure of data



it only for highly demanded privacy devices thanks to the advancement of technology. Figure 8 shows the number of unintentional disclosure of data in our scenario. In our scenario, IoT devices send their data to the cloud server through 42 middle nodes. Among those nodes, 20 ones are malicious nodes, so that they are trying to eavesdrop on the passing data while five nodes inject false data to deviate the data. In general, out of 20 malicious nodes, two ones were able to penetrate the homomorphic encrypted data and find the original data and deviate them.

A simulation of the IoT environment has been run to test our proposed privacy architecture. In addition, CPU consumption should be compared to similar tasks. For the PPDMIT method, each device sends its data to the cloud using Paillier homomorphic encryption. To detect and remove incorrect data in the early phase, the one-way hash chain method has been used in CC. We have also adopted an algorithm for finding valuable data in the Cloud-based on the Gaussian distribution. As well as this, the Chinese residual theorem is applied to data gathered from IoT hybrid devices. The proposed PPDMIT method is compared with LPDA (Lu et al. 2017) and PPIOT (Jayaraman et al. 2017). In this experiment, we investigated the effects of using the key sizes of Paillier processing capability to evaluate the

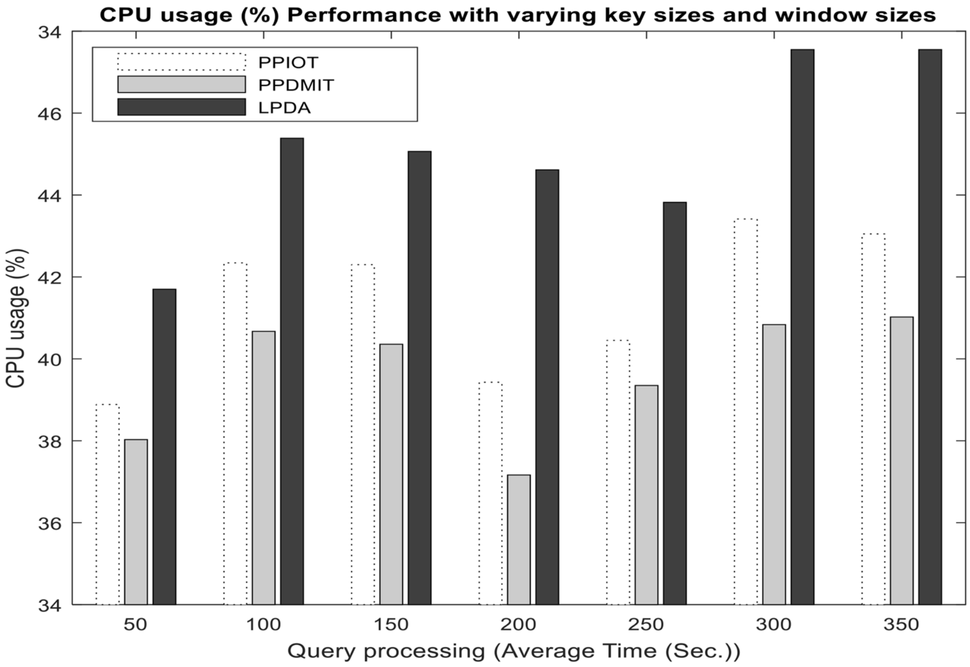


Fig. 9 The details of Computed costs of preserving privacy, extra CPU usage average in LPDA, PPIOT, and PPDMIT

CPU usage average of IoT systems. Since there is no prior key exchange in the proposed scheme, this experiment is relevant.

Consequently, Paillier key combinations of any size are available to the user. A summary of the results can be found in Fig. 9. Observations of the experiments indicate that the key size has a negligible impact on query performance except when 2024 is selected as the key size. When querying sensor data over 9 h, the performance is within 1 s with a 2024 key size. The PPDMIT method is compared to LPDA and PPLOT to show better results for the computed costs of preserving privacy. As can be seen, the proposed method has experienced an 8.096% improvement over LPDA and 6.508% over PPIOT.

7 Conclusion

The quality of life of individuals is expected to be impacted by data analytics based on information from the Internet of Things devices. IoT data aggregation, however, requires careful consideration of security and privacy. A centralized server is usually used to aggregate IoT data. However, it is challenging to coordinate efforts between untrustworthy and sensitive data parties if a distributed approach is used. This paper proposed a method that preserves sensitive data in the Iot Cloud environment while performing data aggregation, PPDMIT. It leveraged four techniques for improving the efficiency: (1) one-way chain for early false detection, (2) homomorphic pallier encryption for encrypting IoT devices' data when they want to send their data to the data aggregator, (3) K-means to find valuable data, and (4) Chinese remainder theorem for data aggregation of IoT devices. We evaluated PPDMIT from the amount of overload from the system perspective. One of the future works is evaluating PPDMIT more comprehensively. Another promising work is applying other privacy-preserving techniques such as the Differential Privacy technique or anonymization technique.

Acknowledgements Special thanks to Islamic Azad Unviersity, Iran. Moreover, this work is supported by Shenzhen Stable Supporting Program (General Project) (No. GXWD20201230155427003-20200821160539001) and Shenzhen Basic Research (General Project) (No. JCYJ20190806142601687).

References

Aïvodji UM, Gambs S, Martin A (2019) IOTFLA : AA secured and privacy-preserving smart home architecture implementing federated learning. In: Proc. - 2019 IEEE Symp. Secur. Priv. Work. SPW 2019, pp. 175–180

Aldeen YAAS, Salleh M, Razzaque MA (2015) A comprehensive review on privacy preserving data mining. Springerplus 4(1):694

Badra M, Zeadally S (2017) Lightweight and efficient privacy-preserving data aggregation approach for the smart grid. Ad Hoc Netw 64:32–40

Beecham (2021) IoT security threat map,Online Report Beecham research. online Rep. http://www.beechamresearch.com/download.aspx?id=43, 2021.

Ding X, Yu Q, Li J, Liu J, Jin H (2013) Distributed anonymization for multiple data providers in a cloud system. In: International Conference on Database Systems for Advanced Applications, pp. 346–360.

Erdos P, Schönheim J (1969) On the set of non pairwise coprime divisors of a number. In: Combinatorial theory and its applications, I (Proc. Colloq., Balatonfüred, 1969), pp. 369–376.

Gheisari M, Wang G, Chen S, Seyfollahi A (2018) A method for privacy-preserving in IoT-SDN integration environment. In: 2018 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Ubiquitous Computing and Communications, Big Data and Cloud Computing, Social Computing and Networking, Sustainable Computing and Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), pp. 895–902.

Gheisari M, Wang G, Chen S (2020) An edge computing-enhanced internet of things framework for privacy-preserving in smart city. Comput Electr Eng 81:106504

Gheisari M et al (2021) OBPP: an ontology-based framework for privacy-preserving in IoT-based smart city. Fut Gen Comput Syst 123:1–13

Guan Z et al (2019) APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT. J Netw Comput Appl 125:82–92

Javadpour A (2019a) Providing a way to create balance between reliability and delays in SDN networks by using the appropriate placement of controllers. Wirel Pers Commun. https://doi.org/10.1007/s11277-019-06773-5

Javadpour A (2019b) Improving resources management in network virtualization by utilizing a software-based network. Wirel Pers Commun 106(2):505–519

Javadpour A, Wang G (2021) cTMvSDN: improving resource management using combination of Markov-process and TDMA in software-defined networking. J Supercomput. https://doi.org/10.1007/s11227-021-03871-9

Javadpour A, Wang G, Rezaei S, Chend S (2018) Power curtailment in cloud environment utilising load balancing machine allocation. In: 2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pp. 1364–1370.

Javadpour A, Wang G, Rezaei S (2020a) Resource management in a peer to peer cloud network for IoT. Wirel Pers Commun. https://doi.org/10.1007/s11277-020-07691-7

Javadpour A, Wang G, Rezaei S, Li K-C (2020b) Detecting straggler MapReduce tasks in big data processing infrastructure by neural network. J Supercomput. https://doi.org/10.1007/s11227-019-03136-6

Jayaraman PP, Yang X, Yavari A, Georgakopoulos D, Yi X (2017) Privacy preserving internet of things: from privacy techniques to a blueprint architecture and efficient implementation. Fut Gen Comput Syst 76:540–549

Jho N-S, Hwang JY, Cheon JH, Kim M-H, Lee DH, Yoo ES (2005) One-way chain based broadcast encryption schemes. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques, pp. 559–574.

Kamakshi P, Babu AV (2012) Automatic detection of sensitive attribute in PPDM. IEEE Int Conf Comput Intell Comput Res 2012:1–5

Khan BUI, Olanrewaju RF, Anwar F, Mir RN, Najeeb AR (2019) A critical insight into the effectiveness of research methods evolved to secure IoT ecosystem. Int J Inf Comput Secur 11(4–5):332–354

Lai J, Li Y, Deng RH, Weng J, Guan C, Yan Q (2014) Towards semantically secure outsourcing of association rule mining on categorical data. Inf Sci (NY) 267:267–286

Liu Q, Wang G, Li F, Yang S, Wu J (2016) Preserving privacy with probabilistic indistinguishability in weighted social networks. IEEE Trans Parallel Distrib Syst 28(5):1417–1429

Lu R, Heung K, Lashkari AH, Ghorbani AA (2017) A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT. IEEE Access 5:3302–3312

Martonosi M (2016) Keynotes: internet of things: history and hype, technology and policy. In: 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), 2016, pp. 1–2.

Melchor CA, Gaborit P (2008) A fast private information retrieval protocol. IEEE Int Symp Inform Theory 2008:1848–1852

Mendel F, Rechberger C, Schläffer M (2009) MD5 is weaker than weak: attacks on concatenated combiners. In: International Conference on the Theory and Application of Cryptology and Information Security, pp. 144–161

Mirmohseni SM, Tang C, Javadpour A (2020) Using markov learning utilization model for resource allocation in cloud of thing network. Wirel Pers Commun. https://doi.org/10.1007/s11277-020-07591-w

Mukkamala R, Ashok VG (2011) Fuzzy-based methods for privacy-preserving data mining. Eighth Int Conf Inform Technol New Gen 2011:348–353

Rachels J (2017) Why privacy is important. In: Privacy, Routledge, pp. 11–21.

Raju R, Komalavalli R, Kesavakumar V (2009) Privacy maintenance collaborative data mining-a practical approach. Second Int Conf Emerg Trends Eng Technol 2009:307–311

Ruj S, Nayak A (2013) A decentralized security framework for data aggregation and access control in smart grids. IEEE Trans Smart Grid 4(1):196–205

Tassa T (2013) Secure mining of association rules in horizontally distributed databases. IEEE Trans Knowl Data Eng 26(4):970–983

Wang Q, Mu H (2021) Privacy-Preserving and Lightweight Selective Aggregation with Fault-Tolerance for Edge Computing-Enhanced IoT. Sensors 21(16):5369

Wang J, Wu L, Zeadally S, Khan MK, He D (2021) Privacy-preserving data aggregation against malicious data mining attack for iot-enabled smart grid. ACM Trans Sen Netw. https://doi.org/10.1145/3440249

Zhang X, Liu C, Nepal S, Yang C, Dou W, Chen J (2013) Combining top-down and bottom-up: scalable sub-tree anonymization over big data using MapReduce on cloud. In: 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, pp. 501–508.

Zhang J, Zhao Y, Wu J, Chen B (2020) LVPDA: a lightweight and verifiable privacy-preserving data aggregation scheme for edge-enabled IoT. IEEE Internet Things J 7(5):4016–4027

Zheng P, Huang J (2013) An efficient image homomorphic encryption scheme with small ciphertext expansion. In: Proceedings of the 21st ACM international conference on Multimedia, pp. 803–812.

Zhu H, Meng X, Kollios G (2014) Privacy preserving similarity evaluation of time series data. EDBT 2014:499–510

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.