Review of Computer Engineering Research
2023 Vol. 10, No. 3, pp. 83-95
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v10i3.3494

A secure and effective data aggregation in WSN for improved security and data privacy

S. Suma Christal Mary[1+]
K. Murugeswari[2]
S. Jyothi Shri[3]
N. Senthamilarasi[4]

[1,4]*Department of Information Technology, Panimalar Engineering College, India.*
[1]*Email: professorsumachristalmary@gmail.com*
[4]*Email: senthamil.pit@gmail.com*
[2]*Department of Computer Science and Engineering, Panimalar Engineering College, India.*
[2]*Email: easwariram13@gmail.com*
[3]*Department of Computer Science and Engineering, SSE, Saveetha Institute of Medical and Technical Sciences, India.*
[3]*Email: drjothishri@gmail.com*

(+ Corresponding author)

Article History
Received: 12 June 2023
Revised: 9 September 2023
Accepted: 19 September 2023
Published: 6 October 2023

Keywords
Data aggregation
Data integrity
Energy consumption
Energy-efficient
Hybrid encryption
Sensitive information
WSN.

ABSTRACT

The increasing prevalence of internet usage and mobile devices has underscored the critical importance of safeguarding personal data. This is especially important in Wireless Sensor Networks (WSNs), where information typically requires in-network computing and collaborative processing. These computationally demanding approaches are not suitable for resource-constrained WSN nodes. Aggregating data effectively while protecting user data is a major challenge in wireless sensor networks. Many privacies of homomorphism encryption-based WSN data aggregation methods have been created and investigated recently. Since cluster leaders (aggregators) may rapidly combine cypher texts without decryption, communication overhead is reduced, making these data aggregation methods more secure than traditional ones. However, the base station only receives aggregated output, causing issues. Initial limits apply to aggregating functions. If the aggregated output is the sum of sensing data, the base station cannot acquire the maximum value. Second, attaching message digests or signatures to sensory samples does not allow the base station to validate data authenticity. In dangerous places, WSNs must be energy-efficient and private. In this research, we present a data aggregation method known as Energy-Efficient and Privacy-Preserving (E2P2). E2P2 data aggregation utilizes less energy and yields more accurate results. Private data aggregation with increased accuracy and hybrid encryption is presented in this research. The goal is to reduce data transmission and energy use collisions and offset collision-induced loss. Extensive simulations compare E2P2 to earlier approaches. Experimental results show that E2P2 outperforms other algorithms. Good exactness, complexity, and safety are demonstrated by theoretical and simulation results.

Contribution/Originality: We suggest an E2P2 data aggregation approach scheme that can reduce data collision, data loss, and overhead, all of which serve to prolong the lifespan of a WSN. In order to protect sensitive information, we present a hybrid approach that combines data slicing and fabricated pieces. To compromise a node's data privacy, an attacker must first break n links belonging to this node and then accurately identify the bogus fragments. The untrusted aggregator can be defeated by our technique, as long as the aggregator itself is not compromised. We drastically cut the privacy protection protocol's energy and communication requirements compared to existing systems.

1. INTRODUCTION

Wireless sensors are currently being employed in a diverse array of sectors. The main issues with wireless sensors are their reminiscence capacity, control consumption, transmission bandwidth, and transmission inconsistency. Processing and storing data from a wireless sensor network (WSN) require a lot of time and effort. Hence, the sensor network stores all data that needs to be processed independently, while the WSNs keep all other data that may be used in the future. A decentralized wireless network is the only one that can provide the amount of RAM that data storage requires. By storing vast amounts of data, it is also important to protect its privacy, security, and veracity. Here, we analyze distributed data storage using the lens of the three security aspects provided by wireless sensor networks. Finally, in a wireless network, a sensor's energy consumption is mostly for data processing rather than data storage.

Several battery-operated nodes with data pre-processing and radio communication capabilities make up a WSN. By allowing sensor nodes to be moved around, dynamic WSNs are more focused and easier to manage than their static counterparts. Hence, dynamic WSNs are often adopted rapidly in applications such as process monitoring, mobility streaming, vehicle status checks, and dairy cattle health monitoring. When it comes to many of the most essential dynamic WSN applications, privacy in WNS is one of the most significant obstacles to overcome [1-3]. As a result, dynamic WSNs have to meet mandatory security requirements.

Upon obtaining data, sensor nodes transmit it to a sink node located elsewhere in the network. Chan, et al. [4] and Yang, et al. [5] have added intrusion detection in cases where some sensors have been compromised. These measures are non-intrusive and are designed to protect users' privacy. And then, after establishing secure communication lines via cryptography, a few constructive privacy-preserving strategies are offered. Eschenauer and Gligor [6] projected the first key pre-distribution system, and subsequent works documented several improved key distribution schemes [7-9]. This pre-distribution of keys can be used to build a safe hop-by-hop data aggregation technique. A straightforward and efficient application of encryption in data aggregation. Yet the key encryption and decryption processes must be carried out locally on each node. As a result, the expense of gathering all that data is substantial. Homomorphic encryption was introduced to build an end-to-end secure data aggregation technique, which allowed for efficiency in privacy-preserving data aggregation. With this method, it is possible to conduct mathematical operations on the ciphertext itself. It is important to remember that key distribution techniques can protect information from being exposed to attackers from outside the network. However, in a stricter scenario, assurances of in-network secrecy may be required. This means that nodes in the network, including the parent node and surrounding nodes, should not have access to any individual's sensitive data. Methods for dealing with these problems are discussed in He, et al. [10], Yang, et al. [11], and Groat, et al. [12].

As a fundamental method in WSNs, data aggregation [13] is commonly employed to combat the problem of excessive power usage. The aggregator can then communicate the total values it calculated from the child sensor nodes to the main aggregator. Redundancy and information synthesis will reduce the load on the network and power needs. Anybody with a suitable wireless receiver can, however, listen in on and potentially intercept communications between sensor nodes during the data aggregation process. To communicate with high-powered computers, the attacker may resort to unethical tactics [14-16]. Theft of data and other forms of illegal involvement can do significant damage to the network and could even bring about its complete paralysis.

Since both economy and confidentiality are crucial factors to consider when building data aggregation algorithms, this research aims to examine security vulnerabilities in data aggregating techniques and provide a novel optimal approach. The nodes' specific energy consumptions, which are used for functions like processing commands, CPU operations, send/receive activities, etc., are what keep the network alive and well. Optimizing some variables, we suggest an E2P2 data aggregation approach scheme that can reduce data collision, data loss, and overhead, all of which serve to prolong the lifespan of a WSN. In order to protect sensitive information, we present a hybrid approach that combines data slicing and fabricated pieces. To compromise a node's data privacy, an attacker must first break n

links belonging to this node and then accurately identify the bogus fragments. The untrusted aggregator can be defeated by our technique, as long as the aggregator itself is not compromised. We drastically cut the privacy protection protocol's energy and communication requirements compared to existing systems.

The subsequent section of this paper delineates the organizational structure. The relevant scholarly works are examined in Part 2, while Section 3 presents initial findings and models. Section 4 offers specific instances of the E2P2 algorithm that we put out. The fifth section presents the outcomes and evaluations of our imitations. The study's concise summary and an explanation of our future objectives are both included in the following section.

2. RELATED WORKS

Many privacy-preserving aggregation techniques exist, and they can be used to protect individuals' anonymity in a variety of contexts. Algorithms that protect users' anonymity are discussed here:

Several methods have been demonstrated in the literature for securely protecting sensitive data at its source. Randomization is a common technique. The randomization approach is a strategy for protecting privacy by introducing randomness into the data. The purpose of this is to hide the information contained in record attributes. There is enough extra noise in the data that it's impossible to get at the original numbers. The values of the data are disrupted, so methods are developed to create aggregated distributions. Afterward, data mining methods can be created to operate on these summative distributions. Historically, randomization has been used to skew results from probability distribution-based approaches like surveys. There are primarily two types of privacy-protecting methods in use. One method incorporates data perturbation techniques, in which a predetermined distribution is superimposed on previously secret information. The aggregate result can be reconstructed from the distribution of the random perturbation. Another method employs the use of random data to hide confidential information. Nevertheless, the problem with data perturbation techniques is that they do not provide reliable aggregate outcomes.

Many useful solutions have been presented to address the problem of connectivity in WSNs. In Xu, et al. [17], an SDA technique based on a grid structure is proposed. A cell's nodes are all capable of instantaneous two-way communication in the adjoining cells since the entire network was partitioned into a set of nonoverlapping virtual cells of a suitable size to guarantee this. Clusters of nodes were formed in accordance with Luo, et al. [18], with one node in each group serving as a gateway to the network. With this method, the technique suggested in Luo, et al. [18] not only guarantees node connectivity but also increases the convergence speed of the network. The grid topology ensures network connectivity, but at the cost of increased computing difficulty as compared to the tree or cluster topologies.

Without a reliable technique to identify interference faults, a receiver node may combine the distorted signals it receives into a single aggregate before sending it on to the root node. There will be widespread corruption of network signals as a result. In applications where batteries can be recharged within a set amount of time, node/link problems in WSN caused by energy drain-out might be regarded as temporary. As a result, even the functional nodes in the network must be down during this transition. For the healthy nodes to remain operational during this transition period, the method of deploying new redundant nodes as a backup to problematic nodes may not be practical. Instead, a method that reconfigures the network by identifying new routes to the BS is a practical and cost-effective alternative. Any failures in the nodes serving as intermediaries will not affect the functionality of the healthy sensor nodes because of this reconfiguration.

Castelluccia, et al. [19] suggested a straightforward and unbreakable dynamically homomorphic stream cryptography that uses slightly more capacity than the hop-by-hop aggregation method. In order to conceal detection and event reporting, Girao, et al. [20] introduced an end-to-end method, which is often more power-efficient than hop-by-hop encryption due to its narrower emphasis on the attacker model [20]. Keeping sensor data private without interfering with additive data aggregation is the goal of a family of techniques presented by Feng, et al. [21]. All of the aforementioned homomorphic encryption methods rely on a symmetric key. These methods' level of security is

proportional to the length of the key used. The intractability of the algorithms is crucial to the safety of the asymmetrical secret key systems. It is for this reason that asymmetric secret-key methods are developed.

Boneh, et al. [22] proposed a homomorphic cryptographically secure election system. Mykletun, et al. [23] investigated the applicability of additively homomorphic public-key data encryption for certain types of wireless sensor systems; their findings helped them select the optimum public key scheme for a certain topology and set of conditions. Ecological fingerprinting may be stored permanently and compactly in encrypted data storage, as proposed by Girao, et al. [24]. In order to make secure data aggregation possible on commercially available sensor systems, Bahi, et al. [25] proposed an end-to-end encrypted data aggregation method to reduce calculation and transmission overhead. For the purpose of protecting data integrity and improving efficiency, Ozdemir and Xiao [26] suggested a new method for aggregating hierarchical hidden data using a security homomorphic data aggregation approach. Lin, et al. [27] hypothesized a fresh approach to securely compiling confidential information. Zhou, et al. [28] organized Secure-Enhanced Data Aggregation, which offers the maximum security for aggregated data of all current asymmetric approaches.

Cluster-Based Privacy Protection Data Aggregation (C-PPDA) was predicted by Dapeng, et al. [29] and by Fang, et al. [30] as a revolutionary energy-efficient Secure Data Aggregation Algorithm (SDAA). Karampour, et al. [31], Kong, et al. [32], Liu, et al. [33], and Liu, et al. [34] have all proposed protocols for data aggregation in the smart grid. In the future, secure multi-party computation [14-16, 35] may also be useful for data aggregation. These publications have contributed to the reduction of traffic congestion; however, further efforts are necessary to enhance their compatibility with users who have more rigorous privacy demands.

3. PRELIMINARIES

The network architecture, attack model, and key management method are presented first as background information to help set the stage for the introduction.

3.1. Network Model

For the purpose of this paper, we wide-range the data aggregation in the system by utilizing the aggregation tree. The network includes a single base station in addition to N individual nodes. These sensor nodes are equipped with the capabilities of sensing, calculating data, and transmitting the information. A non-leaf node will receive data from its child nodes as part of the process of data aggregation. These non-leaf nodes will then combine the data they have received with the data they have collected on their own. The sole tasks that leaf nodes are responsible for are data collection and transmission. The accumulated information will be relayed to the hub via the aggregation tree at some point in the future.

Sensor nodes collect information about their surroundings and relay it to the home base using a number of different data aggregation methods. We incorporate an additive aggregation function into our model. Many more aggregation functions can be derived from the additive aggregation function, making it one of the most fundamental aggregation functions. Standard definitions of data aggregation functions include the following:

$$z(t) = h\{k_1(t), k_2(t) \ldots\ldots\ldots\ldots\ldots k_m(t)\} \quad (1)$$

Where $k_1(t)$ are the data collected from the sensor nodes.

3.2. Attack Model

An adversary can employ a variety of strategies to get around data privacy protections; the following is a synopsis of the opponent's hostile behavior: Attacks based on the study of network traffic. Because wireless communication is used in WSNs, it is possible for adversaries to intercept data while it is being transmitted. We can categorize the attacks into two different sorts, namely global attacks and local attacks, based on the range of the attacks. Attacks

that target congestion control on a global scale are the core interest of our research because they are the most common and straightforward method.

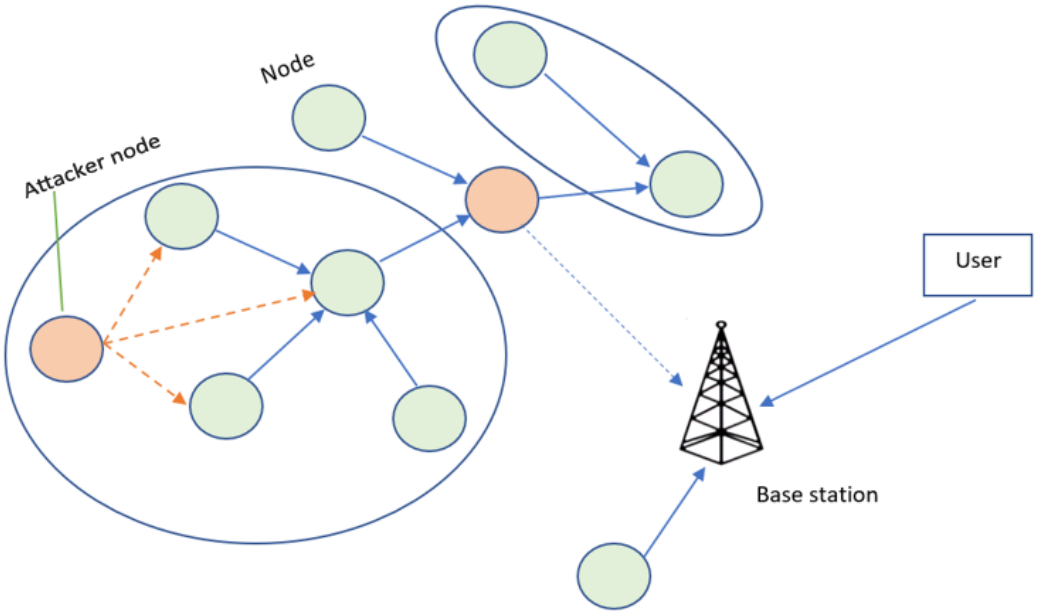

Figure 1. Attack model in WSN sample.

A network assault, known as a passive attack, is one in which a system is checked and maybe sometimes inspected for open ports and vulnerabilities. Passive attacks are a type of network attack. The objective of a passive attack is to acquire knowledge; this type of assault does not entail any action that is taken directly against the target. When a hacker launches an active assault on a network, he or she is trying to change information that is already stored on or in transit to the target. A wide variety of active attacks exist, each with its own advantages and disadvantages. Yet, the threat actor is always accountable for doing anything with the information stored in the system or the gadgets on which the data is kept. A node that has been compromised. Attackers, as shown in Figure 1, only need to compromise a single node to obtain all of the data and keys it contains. The data that an attacker obtains from other nodes can be encrypted, but keys can let them decrypt it. In addition to this, attackers can coordinate their activities on select compromised nodes in order to conclude the primitive data of their neighbor.

3.3. Key Distribution

The data in some messages is typically encrypted before transmission to minimize monitoring by malicious parties. Our model will make use of a proposed random key distribution technique, and we will briefly go over it below. The first step is to build a large key pool containing M keys and their associated identities. By sending discovery messages, sensor nodes in WSN can determine which neighbor share a common key with themselves based on the m keys they arbitrarily select from the key pool. When two nodes in close proximity share a key, they form a secure connection. A link formed by a subset of nodes can serve as a secure connection if they are unable to stake a key. For the aforementioned randomly-generated key distribution system, the likelihood that any given pair of nodes share a key is,

$$N_c = \frac{((M-m)!)^2}{(M-2m)!M!} \quad (2)$$

In addition, for every given key, the likelihood of any other node being able to decode the encrypted message is,

$$N_o = \frac{m}{M} \quad (3)$$

In order to produce an elliptic curve (E) of order $m = p_1 p_2 p_3$, the base station must first create three prime $(p_1, p_2, and\ p_3)$. Thereafter, degree Base Station groups of points $\{A_1, B_2, and\ C_3\}_d$ are picked from E, in the order of m, rendering to the degree of BS which is defined as degree d.

The following formula yields three additional points for each group g:

$$P_g = p_1 p_2 A_1, \quad (4)$$

$$Q_g = p_2 p_3 B_1, \quad (5)$$

$$R_g = p_3 p_1 C_1, \quad (6)$$

P_1 is used to encrypt the merged data, $Q_1$to keep track of the cluster count, and R_1 to mix the encrypted result and fortify data security. A set of keys are then given to the BS. The private key is (p_1, p_2) and the public key is (m, P_g, Q_g, R_g, E). Cluster node masters are given access to the public key, while the BS holds onto the private one.

Once the BS has generated the key, the Cluster head (C) key generation process can commence at each cluster head. For instance, $C(k)$ produces an elliptic curve $(E(k))$ and the three primes $({p_1}^k, {p_2}^k, and\ {p_3}^k)$. in that order. $E(i)$ can be written as,

$$m(i) = {p_1}^k {p_2}^k {p_3}^k \quad (7)$$

Points are chosen from $E^{(i)}$ in the sequence $m^{(i)}$ and grouped into degree $C\_d(i)$ sets by selecting the degree $C\_d(i)$ sets $\{{N_{i1}}^{(k)}, {N_{i2}}^{(k)}, and\ {N_{i3}}^{(k)}\}$ degree $C\_d(i)$. The following formula yields three additional points for each group i:

$${U_i}^k = {p_1}^{(k)} {p_2}^{(k)} {N_{i1}}^{(k)} \quad (8)$$

$${V_i}^k = {p_2}^{(k)} {p_3}^{(k)} {N_{i2}}^{(k)} \quad (9)$$

$${W_i}^k = {p_3}^{(k)} {p_1}^{(k)} {N_{i3}}^{(k)} \quad (10)$$

Here, we use ${U_i}^k$ to encrypt the merged data, ${V_i}^k$ to keep track of the cluster number, and ${W_i}^k$ to mix the encrypted result and strengthen data security.

4. MATERIALS AND METHODS

According to the already available literature, the first step is to build a network display in which the nodes perform the clustering procedure, then to split each node into many nodes, and finally to randomly select a cluster agent for each of the local networks. After that, the central processing units of the individual nodes produce routes and hand out keys. The data is encrypted using the base station's public key at each node. A hash value and time stamp are generated from the encrypted data. The network agent receives one-of-a-kind information from each cluster's nodes. The next step is to compile everything at the cluster agent, compare it to the hash value, and then either use the information if it passes muster or throw it away. In the aggregating phase, all the data is combined before being transmitted to the entry point. Each node in the cluster sends its data to a centralized server, which uses a master key to decrypt and verify the information.

In this section, we will begin by presenting the fundamental concept of E2P2, and then we will discuss the E2P2 system in its entirety. It begins with the phase of tree creation and optimization, then moves on to the phase of slicing and mixing, and finally the phase of aggregation, as shown in Figure 2.

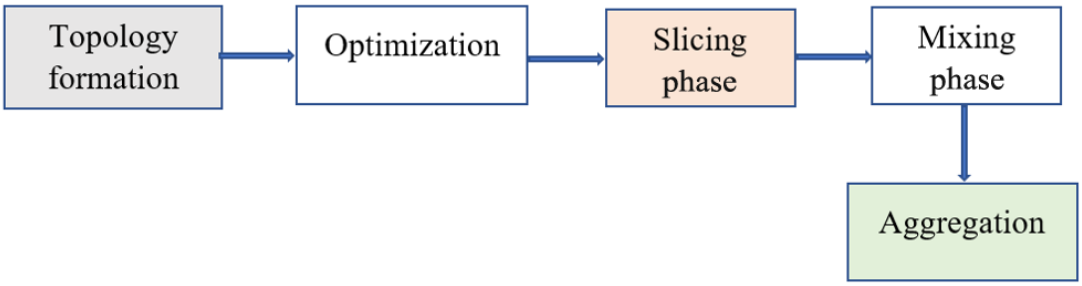

Figure 2. fundamental concept of E2P2.

4.1. Energy-Efficient and Privacy-Preserving (E2P2) Data Aggregation

In this piece, we propose an E2P2 that achieves both low energy consumption and high levels of security. There are three stages to this process. We have the topology-building and optimization stages: The result of our effort to generate a regular tree topology is depicted in Figure 3. Each non-leaf node then chooses one of its leaf child nodes to act as the "chain head" in order to create a chain. As soon as the topology is established, only the leaf nodes begin to partition their information.

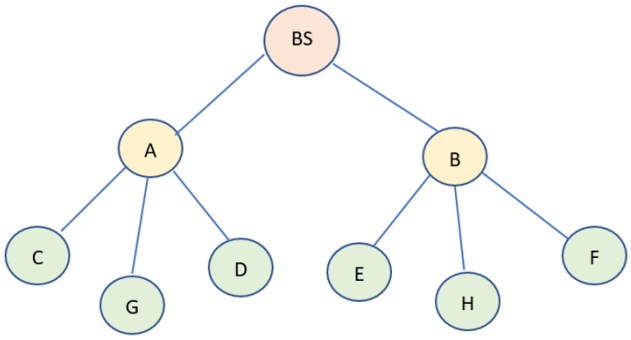

Figure 3. Constructed tree topology.

At the next step, the leaf nodes encrypt the data and send it along to their closest neighbor. Each node will randomly shuffle the bits of data it has received once a specific amount of time has passed. Data is aggregated at each node, and that total is then forwarded to the node that spawned it. In due time, all reports will reach headquarters.

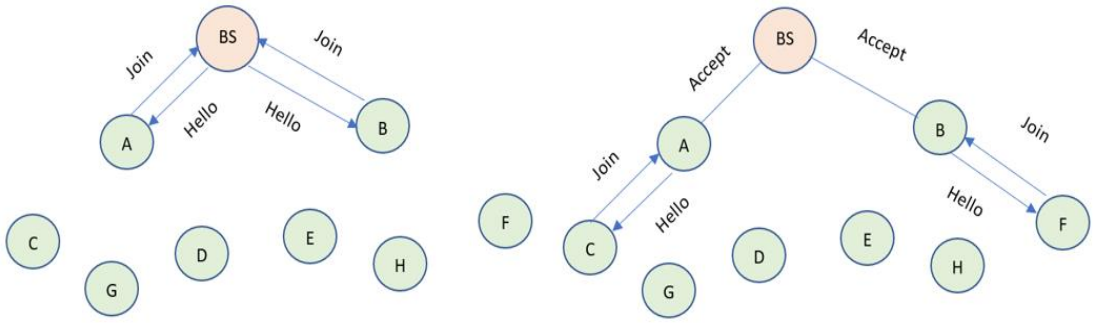

Figure 4. Design of energy-efficient and privacy-preserving (E2P2) data aggregation.

All of the parents send out "Hello" messages to their surrounding nodes, and the orphans can only respond with a "Join" message, as shown in Figure 4. In response to the "Join" message, a parent node will send back an "Accept" message. A node will only respond to the first "Hello" message it gets if it receives two. The tree topology is established once all of the nodes have located their parents. For the purpose of keeping track of its children's IDs, each parent node generates a list. After then, every child node that is not a leaf node will notify its parents that it is not a leaf node by sending a "NOT LEAF" message. Each parent node creates a list of its leaf child nodes by considering all of the children that did not respond with the "NOT LEAF" message to be leaf nodes.

To protect user privacy, leaf nodes randomly slice their data into chunks of size $I_r(2 \leq I_r)$. In the meantime, encrypted data packets will be sent to each node in the network. As a result of data slicing and mixing, nodes will combine their raw data with the bits and bobs they get. No information about the primitives is transmitted to the nodes that contain them. Only if all $I_r - 1$ out-degree links and m in-degree links to a node are hacked will the node's primitive data be revealed. The value of I_r, meanwhile, is produced at random. Unless the maximum number of out-degree linkages is destroyed, attackers cannot be certain that they have successfully broken the defence. One can calculate the likelihood of a leaf node disclosing its own private information as follows:

$$P_f^M = \beta * P_{I_{r-1}} * P_{r+1} \quad (11)$$

β symbolising the odds of identifying a node's right component count.

$$\beta = \frac{1}{r_{max}-2} \quad (12)$$

One subset of non-leaf nodes aggregates data from one child node, whereas another subset combines data from neighboring nodes with their own. Each child node will receive a report of the totals. The equation is,

$$P_f^{M1} = P_1 * P_{r+1} \quad (13)$$

P_1 is the chance that a node's child node will leak some of its data, and P_{r+1} is the chance that any node in its in-degree graph will leak any of its data. The formula for a different class of non-leaf nodes can be deduced using the one shown above.

$$P_f^{M2} = P_c * P_{r+1} \quad (14)$$

5. NETWORK SIMULATION

In this part, a wireless sensor network consisting of 900 nodes is taken into consideration. The nodes in this network are spread out in a random pattern across 500×500 areas. There is a total of $0.5J$ of energy in each node. Within the context of the simulation, we make use of the TAG system, which is a common data aggregation approach. With the use of simulation, we investigate the capabilities of the E2P2 model with regard to the degree distribution, the efficacy of privacy - preserving, the network bandwidth, and the lifetime. In terms of these achievements, comparisons will be made between the E2P2 and BPDA models, the SMART model, and the ESPART model.

6. PERFORMANCE EVALUATION

Here, we assess the confidentiality of the data aggregation methods discussed in this research. Here, we take a look at how well our schemes protect users' privacy, how efficiently they process data, and how accurately they aggregate it all. As a starting point, we utilize TAG, a common data aggregation approach. Data privacy is not protected because it was not a design focus when creating TAG. For the sole purpose of comparing it to our recommended strategies for efficiency and aggregate accuracy, we employ it. First, we define the privacy metric that will be used to measure the efficacy of privacy protection mechanisms. Because of the potential for eavesdropping and collusion, subtle data from individual sensor nodes may be compromised. One's privacy might therefore be compromised in two ways: Once an unwanted node acquires a transmission key, it may read encrypted sensor information. Our key delivery scheme reduces the likelihood that the listener will obtain the shared secret required for communicating between S and one of its neighbours. A few of S's neighbours have conspired to steal sensitive information.

If sensor nodes inside a cluster communicate with one another, only members of that cluster will have access to the sensitive information. One node in a cluster of size k must communicate with the other nodes in the cluster using $k-1$ encrypted messages. A node can only access a member's secret information if it has all of that member's keys $(k-1)$. If not, the confidential information cannot be shared. The $P(m)$ is calculated as,

$$P(m) = \sum_{l=q_c}^{e_{max}} p(n=h)\,(1-(1-s^{l-1})^l) \quad (15)$$

There needs to be some kind of metric used to measure how effective various privacy-preserving strategies are. Many other studies have offered a method that can be summarized as follows. The first premise is that P_{oh} represents the likelihood of eavesdropping at any given node. Indicated by this value is the possibility that every two nodes are in cahoots with one another P_c. In addition, it is assumed in this approach that the two probabilities are equal to one another. In that case, the equation would look like this,

$$P_{oh} = P_c = s \quad (16)$$

Then, for a specific s, the likelihood that node n's secret information is leaked is,

$$P(s) = s^{l-1} \sum_{l=q_c}^{e_{max}} p(I\_d = h)\ .s^l \quad (17)$$

The outdegree of the node is $I-1$, which makes sense. Hence, $P(s)$ can be written in a generic form as,

$$P(s) = \sum_{l=q_c}^{e_{max}} p(d = h) \ .s^l \qquad (18)$$

Rather than focusing on the privacy of a single node, this analysis takes into account the privacy of the entire network. Table 1 shows that Network1, Network2 and Network3 are the two distinct networks. There are 8 vertices in each one. In Network 1 and 3, the nodes' degrees vary widely, however in Network 2, they're all 2.

Table 1. Degree distribution of sample networks.

Nodes	Degree		
	Network 1	Network 2	Network 3
A	3	2	3
B	2	2	2
C	4	2	2
D	3	2	2
E	4	2	4
F	2	2	3
G	4	2	2
H	2	2	4

From Table 1,

$$P_{network\ 1}(s) = \sum_{l=q_c}^{e_{max}} p(d = h) \ .s^l \qquad (19)$$

$$= \frac{3}{8}s^2 + \frac{2}{8}s^3 + \frac{4}{8}s^4$$

$$P_{network\ 1}(s) = \sum_{l=q_c}^{e_{max}} p(d = h) \ .s^l \ = s^2 \qquad (20)$$

$$P_{network\ 3}(s) = \frac{4}{8}s^2 + \frac{2}{8}s^3 + \frac{2}{8}s^4 \qquad (21)$$

In this subsection, we consider a node to have sufficiently safeguarded user privacy if it has a degree of 2. Figure 5 illustrates the degree distribution among the various models. TAG is a data aggregation technique that does not take privacy considerations into account and serves as the foundation for the other models. According to the TAG model, the lowest possible degree is 1, and the highest possible degree is 9. Yet, 70 percent of the nodes in the network have the lowest possible degree. Upon the completion of the privacy-protecting steps, each of the systems raises the lowest degree to 2, and they also raise the supreme degree.

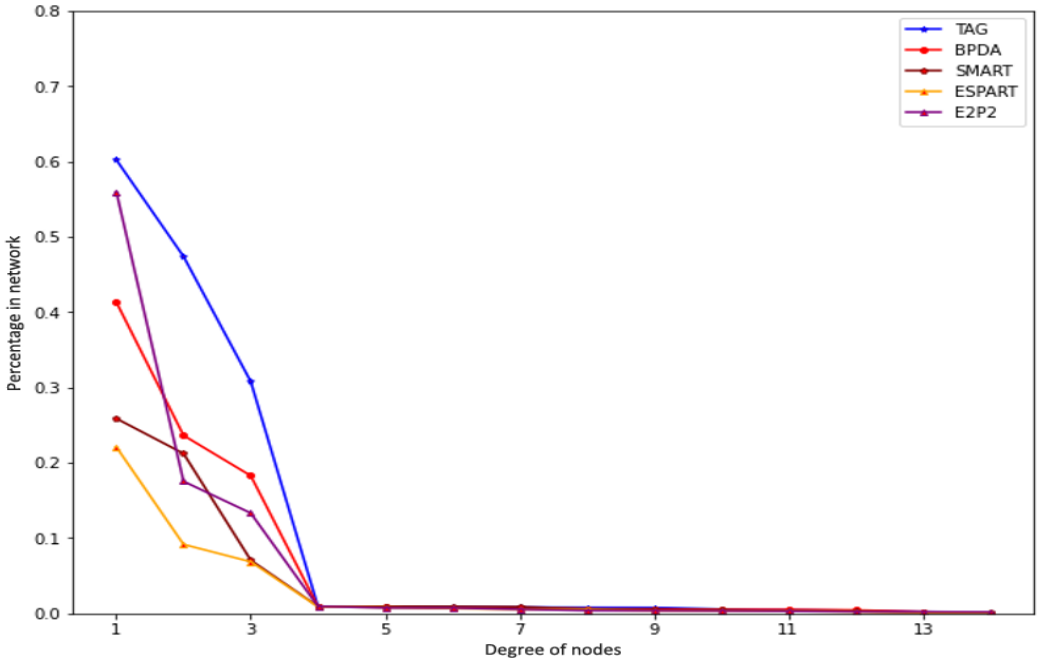

Figure 5. Degree of distribution in different nodes.

Only the BPDA brings the maximum degree up from 8 to 11, and this occurs across BPDA models. The ESPART prototypical and the EBPDA model are identical in that the maximum degree can be raised to 10. In the meantime, the SMART model raises the highest degree possible to 14. Because the maximum degree is being increased, some nodes that do not require having their privacy kept are being preserved anyway. This is the primary factor that contributes to the ineffectiveness of privacy protection measures.

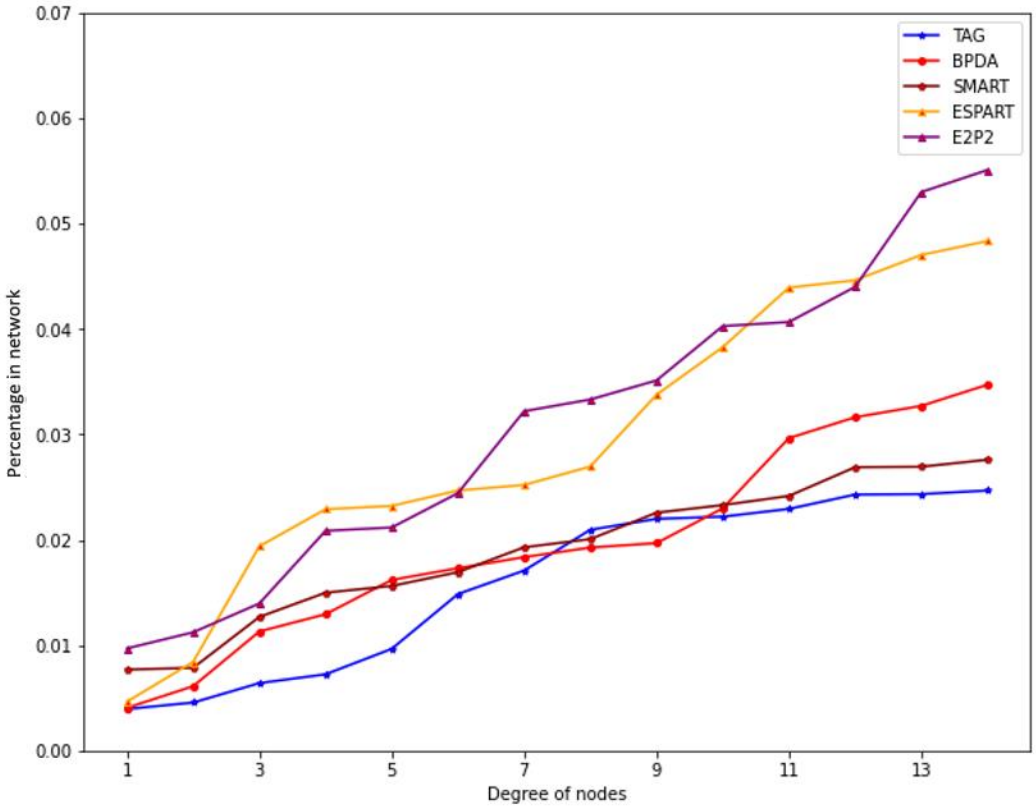

Figure 6. The percentage of nodes that are visible in various algorithms.

As can be seen in Figure 6, the percentage of models whose nodes are visible varies greatly. The unprotected likelihood of nodes in E2P2 models is higher than in BPDA, SMART, and ESPART models Figure 6. This is because this evaluation method looks at the whole WSN. The exposure probability of a model tends to decrease as the sum of its degrees increases. While certain nodes have high privacy protection to a rather high degree after the operation, the BPDA, SMART, and ESPART models appear to have better capability in privacy conservation since they spend considerably more on idleness. Models built using E2P2 take redundancy into account by focusing algorithmic action on the edges with the lowest degrees. Despite the fact that their chances of exposure are greater than average, they are still maintained at a uniform rate.

Figure 7 shows that, as a result, when the number of nodes in a WSN remains constant, the compute consumption of a single query is greater than that of E2P2 and SMART. In particular, when only one slicing mechanism (SMART-1) is used, the SMART calculation overhead is lower than the E2P2 overhead. Yet, compared to E2P2, the security level of one-slice SMART may be lower. Hence, if you're looking for a compromise between security and computational complexity, our technique is the way to go.

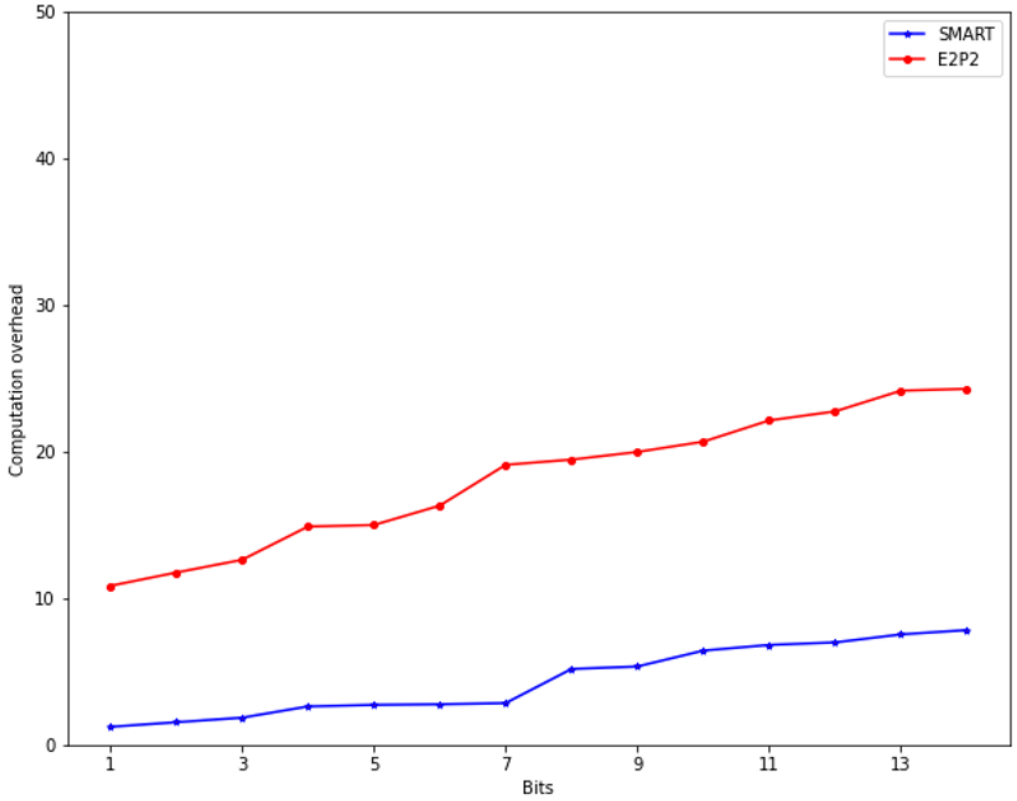

Figure 7. Computation overhead.

7. CONCLUSION

Wireless sensor networks benefit greatly from a type of fundamental and efficient algorithm known as secure data aggregation. In this study, we present E2P2, a brand-new algorithm for aggregation. In the suggested method, a sensor network is structured as an aggregate tree, with links established between the branches. Structural optimization can reduce the number of leaf nodes in the original aggregation tree. Simulation and analysis show that our technique outperforms prior art aggregation methods in efficiency without sacrificing privacy. Two major issues with WSNs are power consumption and data security. Since sensor nodes tend to be installed in high-risk areas, where their data privacy may be more easily compromised than in a cable network, their limited energy may reduce the network's lifespan. The suggested E2P2 method is shown to perform well in terms of accuracy, complexity, and safety in both analytical and simulated settings. From a security standpoint, E2P2 employs the same method as SMART. Hop-by-hop encryption and end-to-end encryption in E2P2 make for an intriguing combination for scheming privacy-preserving data aggregation systems.

In the coming years, our primary focus will be on achieving data aggregation while safeguarding the privacy of individuals, all without relying on any trusted intermediaries. Furthermore, our focus will be directed towards the enhancement of data aggregation methodologies that exhibit superior efficacy, safeguard the privacy of users, and possess the capability to accommodate intricate data formats.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] T. Wang, X. Qin, Y. Ding, L. Liu, and Y. Luo, "Privacy-preserving and energy-efficient continuous data aggregation algorithm in wireless sensor networks," *Wireless Personal Communications*, vol. 98, no. 1, pp. 665-684, 2018. https://doi.org/10.1007/s11277-017-4889-5

[2] B. Murugeshwari, K. Sarukesi, and C. Jayakumar, "An efficient method for knowledge hiding through database extension," presented at the 2010 International Conference on Recent Trends in Information, Telecommunication and Computing. IEEE, 2010.

[3] A. Jenice and D. Hevin, "An energy efficient secure data aggregation in wireless sensor networks," *Research Square*, pp. 1-34, 2021. https://doi.org/10.21203/rs.3.rs-364741/v1

[4] H. Chan, A. Perrig, and D. Song, "Secure hierarchical in-network aggregation in sensor networks," in *Proceedings of the 13th ACM Conference on Computer and Communications Security*, 2006, pp. 278-287.

[5] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A secure hop-by-hop data aggregation protocol for sensor networks," in *Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC '06), Florence, Italy, May 2006*, 2006, pp. 356– 367.

[6] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in *Proceedings of ACM Conference on Computer and Communications Security (CCS '02), Washington, DC, USA, November 2002*, 2002, pp. 41–47.

[7] H. Chan, A. Perrig, and D. X. Song, "Random key predistribution schemes for sensor networks," in *Proceedings of Symposium on Security and Privacy (SP '03), Carnegie Mellon University Pa, USA, May 2003*, 2003, pp. 197–213.

[8] D. Liu, P. Ning, and R. Li, "Establishing pairwise keys in distributed sensor networks," *ACM Transactions on Information and System Security*, vol. 8, no. 1, pp. 41-77, 2005. https://doi.org/10.1145/1053283.1053287

[9] S. Ruj, A. Nayak, and I. Stojmenovic, "Fully secure pairwise and triple key distribution in wireless sensor networks using combinatorial designs," in *Proceedings of the 30th IEEE International Conference on Computer Communications (IEEE INFOCOM '11), Shanghai, China, April 2011*, 2011, pp. 326–330.

[10] W. B. He, X. Liu, and H. Nguyen, "PDA: Privacy-preserving data aggregation in wireless sensor networks," in *Proceedings of the 26th Annual IEEE Conference on Computer Communications (IEEE INFOCOM '07), Anchorage, Alaska, USA, May 2007*, 2007, pp. 2045–2053.

[11] G. Yang, A.-Q. Wang, Z.-Y. Chen, J. Xu, and H.-Y. Wang, "An energy-saving privacy-preserving data aggregation algorithm," *Chinese Journal of Computers*, vol. 34, no. 5, pp. 792-800, 2011. https://doi.org/10.3724/sp.j.1016.2011.00792

[12] M. M. Groat, W. Hey, and S. Forrest, "KIPDA: K-indistinguishable privacy-preserving data aggregation in wireless sensor networks," in *2011 Proceedings IEEE INFOCOM*, 2011: IEEE, pp. 2024-2032.

[13] S. Boubiche, D. E. Boubiche, A. Bilami, and H. Toral-Cruz, "Big data challenges and data aggregation strategies in wireless sensor networks," *IEEE Access*, vol. 6, pp. 20558-20571, 2018. https://doi.org/10.1109/access.2018.2821445

[14] Y. Wang, G. Yang, T. Li, F. Li, Y. Tian, and X. Yu, "Belief and fairness: A secure two-party protocol toward the view of entropy for IoT devices," *Journal of Network and Computer Applications*, vol. 161, p. 102641, 2020. https://doi.org/10.1016/j.jnca.2020.102641

[15] Y. Wang *et al.*, "Incentive compatible and anti-compounding of wealth in proof-of-stake," *Information Sciences*, vol. 530, pp. 85-94, 2020. https://doi.org/10.1016/j.ins.2020.03.098

[16] Y. Wang *et al.*, "Optimal mixed block withholding attacks based on reinforcement learning," *International Journal of Intelligent Systems*, vol. 35, no. 12, pp. 2032-2048, 2020. https://doi.org/10.1002/int.22282

[17] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, 2001, pp. 70-84.

[18] X. Luo, Y. Yan, S. Li, and X. Guan, "Topology control based on optimally rigid graph in wireless sensor networks," *Computer Networks*, vol. 57, no. 4, pp. 1037-1047, 2013. https://doi.org/10.1016/j.comnet.2012.12.002

[19] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*, 2005: IEEE, pp. 109-117.

[20] J. Girao, D. Westhoff, and M. Schneider, "CDA: Concealed data aggregation for reverse multicast traffic in wireless sensor networks," in *Proceedings of the IEEE International Conference on Communications (ICC '05)*, 2005, pp. 3044-3049.

[21] T. Feng, C. Wang, W. Zhang, and L. Ruan, "Confidentiality protection for distributed sensor data aggregation," in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*, 2008: IEEE, pp. 56-60.

[22] D. Boneh, E.-J. Goh, and K. Nissim, "Evaluating 2-dnf formulas on ciphertexts," in *Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005, Cambridge, MA, USA, February 10-12, 2005. Proceedings, Lecture Notes in Computer Science, 2005* 2005, pp. 325–341.

[23] E. Mykletun, J. Girao, and D. Westhoff, "Public key based cryptoschemes for data concealment in wireless sensor networks," presented at the 2006 IEEE International Conference on Communications, 2006.

[24] J. Girao, D. Westhoff, E. Mykletun, and T. Araki, "Tinypeds: Tiny persistent encrypted data storage in asynchronous wireless sensor networks," *Ad Hoc Networks*, vol. 5, no. 7, pp. 1073-1089, 2007. https://doi.org/10.1016/j.adhoc.2006.05.004

[25] J. M. Bahi, C. Guyeux, and A. Makhoul, "Efficient and robust secure aggregation of encrypted data in sensor networks," in *2010 Fourth International Conference on Sensor Technologies and Applications*, 2010: IEEE, pp. 472-477.

[26] S. Ozdemir and Y. Xiao, "Integrity protecting hierarchical concealed data aggregation for wireless sensor networks," *Computer Networks*, vol. 55, no. 8, pp. 1735-1746, 2011. https://doi.org/10.1016/j.comnet.2011.01.006

[27] Y.-H. Lin, S.-Y. Chang, and H.-M. Sun, "CDAMA: Concealed data aggregation scheme for multiple applications in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1471-1483, 2012. https://doi.org/10.1109/tkde.2012.94

[28] Q. Zhou, G. Yang, and L. He, "A secure-enhanced data aggregation based on ECC in wireless sensor networks," *Sensors*, vol. 14, no. 4, pp. 6701-6721, 2014. https://doi.org/10.3390/s140406701

[29] M. Dapeng, W. Chenye, Y. Wu, W. Wei, X. Shichang, and J. Xiaopeng, "Energy-efficient cluster-based privacy data aggregation for wireless sensor networks," *Journal of Tsinghua University (Science and Technology)*, vol. 57, no. 2, pp. 213-219, 2017.

[30] W. Fang, X. Wen, J. Xu, and J. Zhu, "CSDA: A novel cluster-based secure data aggregation scheme for WSNs," *Cluster Computing*, vol. 22, no. S3, pp. 5233-5244, 2019. https://doi.org/10.1007/s10586-017-1195-7

[31] A. Karampour, M. Ashouri-Talouki, and B. T. Ladani, "An efficient privacy-preserving data aggregation scheme in smart grid," presented at the 2019 27th Iranian Conference on Electrical Engineering (ICEE), 2019.

[32] W. Kong, J. Shen, P. Vijayakumar, Y. Cho, and V. Chang, "A practical group blind signature scheme for privacy protection in smart grid," *Journal of Parallel and Distributed Computing*, vol. 136, pp. 29-39, 2020. https://doi.org/10.1016/j.jpdc.2019.09.016

[33] X. Liu, X. Yu, H. Zhu, G. Yang, Y. Wang, and X. Yu, "A game-theoretic approach of mixing different qualities of coins," *International Journal of Intelligent Systems*, vol. 35, no. 12, pp. 1899-1911, 2020. https://doi.org/10.1002/int.22277

[34] X. Liu, J. Yu, X. Zhang, Q. Zhang, and C. Fu, "Energy-efficient privacy-preserving data aggregation protocols based on slicing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, pp. 1-12, 2020. https://doi.org/10.1186/s13638-020-1643-6

[35] Y. Wang, A. Bracciali, G. Yang, T. Li, and X. Yu, "Adversarial behaviours in mixing coins under incomplete information," *Applied Soft Computing*, vol. 96, p. 106605, 2020. https://doi.org/10.1016/j.asoc.2020.106605