Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 272537, 10 pages
http://dx.doi.org/10.1155/2014/272537

Research Article

A Signature-Based Data Security Technique for Energy-Efficient Data Aggregation in Wireless Sensor Networks

Min Yoon, Miyoung Jang, Hyeong-Il Kim, and Jae-Woo Chang

Department of Computer Engineering, Chonbuk National University, Chonju, Chonbuk 561-756, Republic of Korea

Correspondence should be addressed to Jae-Woo Chang; jwchang@jbnu.ac.kr

Received 26 July 2013; Revised 13 December 2013; Accepted 15 December 2013; Published 16 January 2014

Academic Editor: Jong Hyuk Park

Copyright © 2014 Min Yoon et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data aggregation techniques have been widely used in wireless sensor networks (WSNs) to solve the energy constraint problems of sensor nodes. They can conserve the significant amount of energy by reducing data packet transmission costs. However, many data aggregation applications require privacy and integrity protection of the real data while transmitting data from the sensing nodes to a sink node. The existing schemes for supporting both privacy and integrity, that is, iCDPA, and iPDA, suffer from high communication cost, high computation cost, and data propagation delay. To resolve the problems, we propose a signature-based data security technique for protecting sensitive data aggregation in WSNs. To support privacy-preserving data aggregation and integrity checking, our technique makes use of the additive property of complex numbers. Out of two parts of a complex number, the real part is used to hide the sampled data of a sensor node from its neighboring nodes and adversaries, whereas the imaginary part is used for data integrity checking at both data aggregators and the sink node. Through a performance analysis, we prove that our privacy-preserving data aggregation scheme outperforms the existing schemes up to 50% in terms of communication and computation overheads as well as up to 3 times in terms of integrity checking and data propagation delay.

1. Introduction

Wireless sensor networks (WSNs) have been widely studied in ubiquitous computing environment. The WSNs can be applied to various types of applications, such as environment management and military monitoring [1–4]. However, the sensor nodes that form WSNs have resource constraints such as limited power, slow processor, and less memory. For these reasons, it is essential to improve the energy efficiency of sensor nodes (or WSN) in order to enhance the quality of application service [5–10]. The first issue of WSNs is to reduce energy consumption in WSNs. Because the amount of energy consumption for communication is the greatest, it is important to reduce communication overhead. For reducing communication cost, transmitting the required and partially processed data is more meaningful than sending a large amount of raw data. In general, sending raw data causes the energy consumption of sensor nodes because duplicated messages are sent to the same node, called implosion, as well as neighboring nodes receive the duplicated messages if two nodes share the same observing region, called overlapping. In recent years, data aggregation has been actively used to combine data coming from many sensor nodes. An extension of this approach is in-network aggregation which aggregates data progressively as data are passed through the network [11–14]. In-network data aggregation can reduce the number of data transmissions and the number of nodes involved in gathering data from a WSN.

The second issue of WSNs is how to preserve sensitive measurements where data privacy becomes an important aspect from an adversary [15]. In many scenarios, the confidentiality of transported data can be considered critical. For instance, data from sensors might measure patients' health information such as heartbeat and blood pressure details. In addition, a future application might measure household details such as power and water usage, thus computing average trends and making local recommendations. Since sensitive data is transported wirelessly among sensor nodes, it is typically prone to interception and eavesdropping. It is mandatory to maintain the data privacy of sensor nodes even from other trusted participating sensor nodes of the WSNs. As a result, even though private data are overheard and

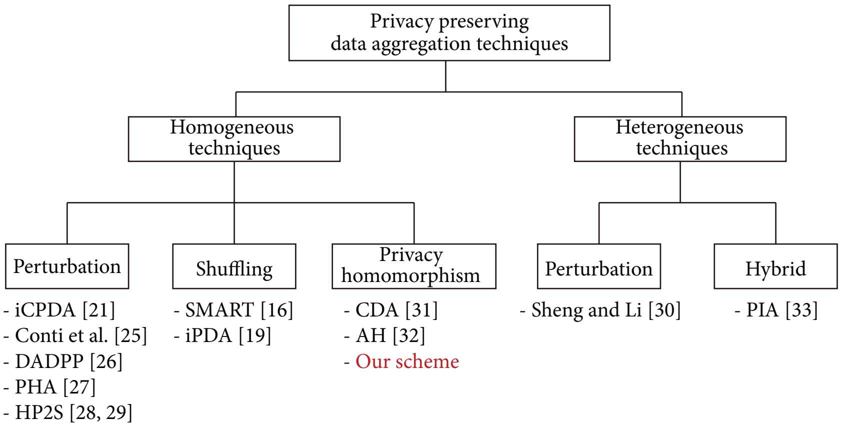


Figure 1: Classification of the privacy preserving-data aggregation techniques for WSNs.

decrypted by adversaries, it is necessary to prevent recovering the sensitive information of a sensor node [16–18].

The last issue of WSNs is data integrity [19–21]. In communication, data integrity is simply defined as maintaining consistency and correctness of messages (message without modification by adversaries). In other words, it is ensured that the received data is not altered in transit either by an adversary or by noise in the data collecting node, that is, sink node. Data pollution due to the noise is an unintentional process and it can be handled by using some existing mechanisms like cyclic redundancy checking (CRC). Hence, the integrity checking due to the unintentional data pollution is out of the scope of this research. On the other hand, the mechanisms like CRC are unable to cope with the intentional data pollution by an adversary because the adversary can generate the same CRC of the source node after modifying the data. As data aggregation result is used for making critical decisions, the aggregation result must be verified before accepting it. For this reason, it is required to design a data protocol for WSNs which can ensure that the aggregated result has not been polluted (manipulation of data by an adversary) on the way to the sink node.

Since data privacy and integrity protection processes consume a significant amount of precious resource (i.e., limited power) of sensor nodes, they shorten the lifetime of the WSNs. Therefore, it is necessary to devise a light-weight technique, which can achieve data privacy and integrity protection efficiently. However, the existing work needs much resource consumption of sensor nodes due to generating unnecessary messages in the network. For this reason, in this paper, we propose a resource-efficient data security technique that can aggregate sensitive data while protecting data integrity in WSNs. Our technique protects from the leak of the sensed data by using the algebraic properties of the complex numbers. Our technique not only ensures that no trend about the sensitive data of a sensor node is released to any other nodes and adversaries, but also can aggregate and hide data for data privacy during transmissions to the data sink. Out of two parts of a complex number, the real part is used to hide the sampled data of a sensor node from its neighboring nodes and adversaries, whereas the imaginary part is used for data integrity. Before transmitting data to a parent node, every sensor node transforms its sampled data into a complex number form. The real part is generated by combining the sampled data with a unique private seed and the imaginary part is generated by appending an imaginary unit to the modified sampled data. Thus, our technique prevents from recovering sensitive information even though private data are overheard and decrypted by adversaries or other trusted participants. For strong data security, our technique can be built on the top of the existing secure communication protocols like [22]. Moreover, our technique can be applied to any type of WSNs regardless of network topology since it is a general approach.

The rest of the paper is organized as follows. In Section 2, we present some related work. Section 3 describes our integrity-protecting sensitive data aggregation technique. Simulation results are shown in Section 4. Along with some future research directions, we finally conclude our work in Section 5.

2. Related Work

In this section, we present related work for privacy-preserving data aggregation schemes. Figure 1 illustrates the classification of the privacy-preserving data aggregation techniques for WSNs. These techniques are broadly categorized into two categories: homogeneous techniques and heterogeneous ones. They are categorized based on the type of nodes in the WSNs, particularly the type of data aggregating nodes (aggregators). The aggregators can either be special (more powerful) nodes or regular sensor nodes. Moreover, the techniques are further divided into five groups: perturbation in homogeneous technique, shuffling, privacy homomorphism, perturbation in heterogeneous, and hybrid. First, the perturbation technique is also known as data customization. In this technique, every sensor node uses encryption key and/or seeds (private or public) generated by randomization techniques [23, 24] in order to hide the sampled data before transmitting them to a parent node. The perturbation in homogeneous technique include iCPDA [21], Conti et al.'s scheme [25], DADPP [26], PHA [27], and HP2S [28, 29], while the perturbation technique in heterogeneous includes Sheng and Li's scheme [30]. Second, in the shuffling technique, every sensor node slices its data into the fixed number (J) of data pieces and sends a data piece to the selected $J-1$ number of neighboring sensor nodes. The remaining one piece of data is kept with it. After that, every sensor node assembles the received data pieces including its own piece of data and sends the assembled data to a parent node. SMART [16] and iPDA [19] belong to the shuffling techniques. Third, the privacy homomorphism technique has a special feature that allows arithmetic operations to be performed on ciphertext without decryption. This technique is fast and resource-efficient for privacy-preserving data aggregation, but it has a limitation that it performs only addition and multiplication operations. Before the sensed data are sent to the aggregators, they are encrypted by using the respective keys of sensor nodes and they are added or multiplied without decryption. The CDA [31], AH scheme [32], and our scheme belong to the privacy homomorphism techniques. Finally, the hybrid technique achieves privacy-preserving data aggregation for WSNs by combining the previous techniques. PIA [33] is only the hybrid technique in this literature.

In the previous section, we addressed three important considerations for WSNs, which are energy consumption, data privacy, and data integrity. However, iPDA and iCPDA are the only works to support both privacy preservation and data integrity for WSNs; we provide the detailed explanation of iPDA and iCPDA in Section 2.1.

2.1. Privacy Preserving Data Aggregation Scheme with Data Integrity. He et al. proposed iPDA [19] and iCPDA [21] schemes for WSNs to support privacy-preserving data aggregation as well as data integrity. In the iPDA scheme, they protect data integrity by designing two node-disjoint aggregation trees rooted at the query server where each node belongs to a single aggregation tree. In this technique, first, every sensor node slices its private data randomly into L pieces and $L-1$ pieces are encrypted and sent to the randomly selected sensor nodes of the aggregation tree keeping one piece at the same sensor node. The same process is independently done for each sensor node using another aggregation tree. Then, all the sensor nodes which received data slices from multiple sensor nodes decrypt the slices using their shared keys and sum the received data slices including their own. After that, each sensor node sends the sum value to its parent from the respective aggregation tree. In the same way, the sum data from another set of sensor nodes are transmitted to the query server through another aggregation tree. In the end, the aggregated data from two node-disjoint aggregation trees reach to the base station where the aggregated data from both aggregation trees are compared. If the difference of the aggregated data from the two aggregation trees does not deviate from the predefined threshold value the query server accepts the aggregation result; otherwise, it rejects the aggregated result by considering it as polluted data. However, there are some shortcomings in the iPDA. First of all, during protecting data privacy it generates high traffics in the WSN. As a result, communication cost is significantly increased in the iPDA. Secondly, all sensor nodes use secret keys to encrypt all of their data slices before sending to their respective $2(L-1)$ number of sensor nodes. So, every sensor node has computation overhead of decrypting all the slices they received before aggregating them.

In the iCPDA, three rounds of interactions are required. Firstly, each node sends a seed to other cluster members. Next, each node hides its sensory data via the received seeds and sends the hidden sensory data to each cluster member. Then, each node adds its own hidden data to the received hidden data and sends the calculated results to its cluster head which calculates the aggregation results via inverse and multiplication of matrix. To enforce data integrity, cluster members check the transmitted aggregated data of the cluster head. There are some disadvantages of iCPDA. Firstly, the communication overhead of iCPDA increases quadratically with the cluster size. Secondly, the computational overhead of CPDA increases quickly with the increase of the cluster size which introduces large matrix, whereas lower cluster size introduces lower privacy-preserving efficacy.

Both iPDA and iCPDA support very weak data integrity checking because if any node modifies its sampled value 30 to 300 and uses the value 300 for aggregation process none of both methods can detect such misbehavior in the network. Hence, in this paper, we propose a new, efficient (in terms of communication overhead and data propagation delay), and general (in terms of supporting network topology) scheme in order to support data privacy and achieve integrity assurance in data aggregation for WSNs. Our scheme is based on the algebraic properties of the complex numbers and it not only ensures that no trend about sensitive data of a sensor node is released to any other nodes and adversaries but also provides data integrity checking of the aggregated value of sensor data.

3. Integrity-Protecting Sensitive Data Aggregation Technique

To overcome the previously mentioned shortcomings of the iPDA and iCPDA, in this section, we propose a new energy-efficient data aggregation scheme for preserving data privacy in WSNs. Our scheme exploits an additive property of complex number to aggregate the sensed data in WSNs. Our assumption is that we only focus on additive aggregation function (SUM), like the iCPDA and iPDA. This is because other aggregation functions, such as average, count, variance, and standard deviation, can be obtained by using the additive aggregation function [34]. In our scheme, out of two parts of a complex number $(a+bi)$, the real part (a) is used to hide the sampled data of a sensor node from its neighboring nodes and adversaries, whereas the imaginary part (bi) is used for data integrity checking at both data aggregators and the sink node. Before transmitting data to a parent node, every sensor node transforms its sampled data into a complex number form. The real part is generated by combining the sampled data with a unique private seed and the imaginary part is generated by appending an imaginary unit to the modified sampled data. For this, the sampled value is first mingled with a private seed and then the result (a) is combined with another real number

Table 1: Real ID of 8 sensor nodes with signature.

SN	Node-ID	2-Byte signature
1	$2^0 = 1$	0000000000000001
2	$2^1 = 2$	0000000000000010
3	$2^2 = 4$	0000000000000100
4	$2^3 = 8$	0000000000001000
5	$2^4 = 16$	0000000000010000
6	$2^5 = 32$	0000000000100000
7	$2^6 = 64$	0000000001000000
8	$2^7 = 128$	0000000010000000

having i (bi) to generate a complex number form ($c = a + bi$). The real number with i (bi) is the absolute difference between the previous sample data and the current sample data of a node. Note that during network deployment, a Master Device (MD) [35] securely provides a unique real number as a seed to every sensor node of the WSNs after establishing a pairwise secret key with them. Since the MD is an offline server, it shares this information only with the query server for future reference. Thus, the seed of each sensor node is private in the network. Data can be aggregated in upper levels during their transmissions to the query server by using the algebraic properties of complex numbers. Our scheme can check the integrity of the aggregated data at both data aggregators and the sink node at the same time.

The proposed privacy and integrity preserving technique is performed through five steps. In the first step, we assign a special type of positive integer 2^n (where $n = 0$ to $Bn \times 8 - 1$, such that Bn is the number of free bytes available in the payload) to every sensor node as node ID. This is because the binary value of every integer of 2^n type has only one high bit (1). In addition, the position of the high bit for all integers of this type is unique. The sink node knows a data contributing sensor node through the signature of Node-ID as shown in Table 1. The Node-ID of a sensor node is used to generate a signature of a fixed length. A signature is a fixed size bit stream of binary numbers for a given integer. Signature of a senor node ID can be generated by using the technique presented in the work [36]. We can determine the length of the signature based on the size of a given WSN. When the size of the WSN increases we can increase the length of the signature up to the Bn bytes. In other words, different size WSNs can have signatures of different lengths. The detail of using signatures has been presented in our previous work [37].

When the network receives an SQL-like query for SUM aggregation function, in the second step, the sampled sensitive data ds of each sensor node is, first, concealed in a by combining with a unique seed (sr) which is a private real number. The seeds can be selected from an integer range (i.e., space between lower bound and upper bound). By increasing the size of the range, we can further increase the level of the data privacy. Hence, our approach can support data privacy feature strongly. To support data integrity, an integer value b—*the difference of the previous sensed value and the current sensed value* of the sensor node—with i is appended to the a by using *genCpxNum*() function to form a complex number $C = a + bi$, where a and b are real numbers called the real part and the imaginary part of the complex number, respectively, as shown in Table 2. Complex numbers can be added, subtracted, multiplied, and divided by formally applying associative, commutative, and distributed laws of algebra. For the first round, the complex number (value of b) is zero. In Table 2, for instance, the reading 17 of node 5 is encrypted into $46 + 3i$. The reading 17 is added to 29, which is a private seed of node 5 and the mask value 46 is calculated. Then, assuming that $3i$ is the difference value of previous reading and current reading of node 5, the $3i$ is appended to the result 46 to get $46 + 3i$ which is a complex number form of the 17 after data customization process. Node 5 includes its signature, that is, 00000101, when it transmits the data as $\langle 00000101, 46 + 3i \rangle$. We assumed that any sensor node cannot be compromised before sending first round data to the sink node. Every source sensor node keeps the original sensed value d of the current round to deduce b in the next round which is updated in each round of data transmission. Next, the source node encrypts the customized data R_1', that is, $R_1 = a + bi$, and the signature of the node by using a secret key Kx, y [22] and transmits the cipher text C_j to its parent. The term Kx, y denotes a pairwise symmetric key shared by nodes x and y, where the node x encrypts data by using a key Kx, y and the node y decrypts the data by using the key Kx, y. In this way, our algorithm converts the sampled data into an encrypted complex number form. Hence, it not only protects the transmitting trend of private data but also does not let neighboring sensor nodes and adversaries to recover sensitive data even though they overheard and decrypted the sensitive data.

In the third step, the parent sensor node (i.e., data aggregator) decrypts the received data by using respective pairwise symmetric keys of its child sensor nodes. For each child node, the parent node computes the difference value (b') of the two real units by using the stored previous data and the received current data of the child node. For the first round, the value of b' is also zero. For this, the parent node always keeps the record of the previously received data from each of the child nodes and it updates the previous data by current one in every round. To support local integrity checking, the parent node first compares just computed difference value with the currently received difference value (imaginary unit) from the child node and then compares the difference value with local threshold δ. If the imaginary unit of the child's current data is equal to the computed difference value and the imaginary unit is not greater than δ, then the parent node accepts the data of the child node. Otherwise, the parent node rejects the data of the child sensor node considering it as polluted data. For example, we assume that the value for δ is set to 2 for local integrity checking. Because a parent node checks the integrity of its' child nodes, node 4 checks the local integrity of the node 8. In Figure 1, since the imaginary part of node 8 is 2, which is less than or equal to δ, node 4 accepts the data of node 8. On the other hand, node 5 will be rejected by its parent node 2 because imaginary part of node 5 is greater than δ. In the same way, the parent node assures the data integrity of child nodes. After that the parent node adds the data of child nodes including its own by using additive property

Table 2: Customized data creation for each node.

SN	Reading (ds)	Real seed (sr)	Mask value (a = ds + sr)	Difference value (bi)	Complex number ($a + bi$)
1	16	40	56	$2i$	$56 + 2i$
2	14	51	65	$0i$	$65 + 0i$
3	19	32	51	i	$51 + i$
4	21	23	44	i	$44 + i$
5	17	29	46	$3i$	$46 + 3i$
6	18	33	51	i	$51 + i$
7	13	39	52	$2i$	$52 + 2i$
8	15	67	82	$2i$	$82 + 2i$

of complex number to produce an intermediate result R'. At the same time, it superimposes signatures (SSig) of the contributed nodes by performing bitwise OR operation on the bit-streams of the node IDs and forwards the encrypted intermediate result "C_r" towards the sink node. Since this approach needs just one bit to carry an ID of a sensor node it is 16 times scalable than the existing work CMT [34] where plaintexts (2-byte each) are used for carrying IDs of sensor nodes by simply concatenating them. Note that different types of application can have different value for the threshold δ. Thus, our algorithm supports local integrity checking which enforces to provide consistent data from child nodes. The above process continues at all nodes of the upper levels of the network until the whole partially aggregated data of the network reach to the sink node.

In the fourth step, when the sink node receives all intermediate result sets C_{rs} (partially aggregated encrypted customized data with superimposed signature) from the 1-hop child nodes, it decrypts them by using respective pairwise symmetric keys and computes the final aggregation SUM_2 from C_{rs}. Since SUM_2 is of complex number form and the sensed data has been concealed in the real unit by using private seeds, identifying the information of the contributed sensor nodes is necessary to deduce actual SUM value. In the last step, the sink node first knows data contributing nodes by checking the high bits (1 s) of the received superimposed signature by performing bitwise AND operation with the prestored signature files or superimposed signature of the Node-IDs of the all nodes of the network. For this, it separates SUM_2 into real unit $\text{SUM}_{2\text{R}}$ and imaginary unit $\text{SUM}_{2\text{IM}}$. Because the sampled data of sensor nodes has been concealed within the real unit, the sink node computes the actual aggregated result SUM by subtracting (an inverse operation of masking, step 2) $\text{SUM}_{1\text{R}}$ (a freshly computed sum value of the private seeds of the contributed source nodes) from $\text{SUM}_{2\text{R}}$. The final result SUM is always accurate and reliable because of the following two reasons. First, a complex number is an algebraic expression and hence the underlying algebra gives the accurate result of the aggregated sensor data. Second, since the private seeds are fixed integer values (i.e., seeds are not random numbers) after collecting data by the sink node a complex number subtracts exactly the same values that have been added to the sensor data during data hiding process by every source node. At the same time, before accepting the SUM, the sink node performs global integrity checking of SUM to assure whether the SUM_2 has been polluted by an adversary in transit or not. For this, like parent nodes, the sink node also computes the difference value (B') of the two real units by using the stored previous data and the received current data from the network. The sink node first compares just computed difference value $B'i$ with the currently received difference value, that is, $\text{SUM}_{2\text{IM}}$, from the network and then compares the difference value ($\text{SUM}_{2\text{IM}}$) with global threshold Δ (for every application, the maximum value for $\Delta = \delta \times N$, where N is the total number of nodes in a network). If the imaginary unit $\text{SUM}_{2\text{IM}}$ of the current data from the network is equal to the just computed difference value $B'i$ and the $\text{SUM}_{2\text{IM}}$ is not larger than Δ, then the sink node accepts the data of the network and returned the actual SUM to the query issuer. Otherwise, the sink node rejects the SUM considering it as forged/polluted data by adversary or other nodes. For example, as shown in Figure 2, we assume that a local integrity threshold per node δ equals to $2i$ and the maximum value for a global threshold (Δ) is calculated as $\Delta = \delta \times N = 2i \times 8 = 16i$. Since a sensor node 5 does not participate in data collection, the global integrity checking value Δ can be computed as $\delta \times N = 2i \times 7 = 14i$. In this scenario, the received data is considered as a consistent one and is accepted by the sink node, (1) because the value computed at the sink node, that is, $9i$, is the same as the one received from the network and (2) the value is less than the global integrity checking value, that is, $9i < 14i$. The overall algorithm that performs sensitive data aggregation and integrity checking is illustrated in Algorithm 1.

4. Performance Evaluation

In this section, we present simulation results of our scheme by comparing it with iPDA and iCPDA schemes in terms of communication overhead and integrity checking. For this, we use TOSSIM [38] simulator running over TinyOS [39] operating system and GCC compiler. We consider 100 sensor nodes distributed randomly in 100 m × 100 m area. As presented in directed diffusion [40], we use such parameters as receiving power dissipation of 395 mW and transmitting power dissipation of 660 mW. Moreover, MATLAB 7.6.0.324 (R2008a) is used to get execution time required for data customization and data aggregation.

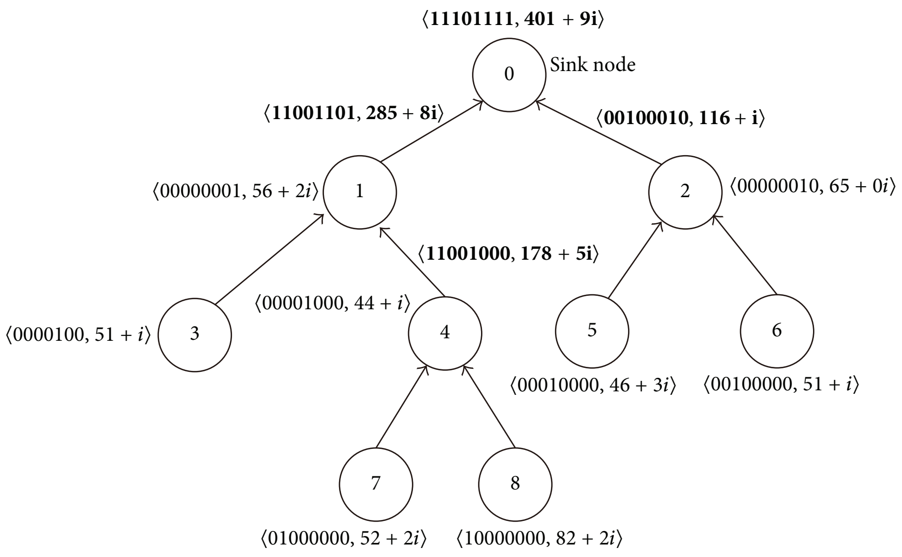


Figure 2: Superimposing signatures and addition of customized sensor readings in a multihop WSN ($\delta = 2$).

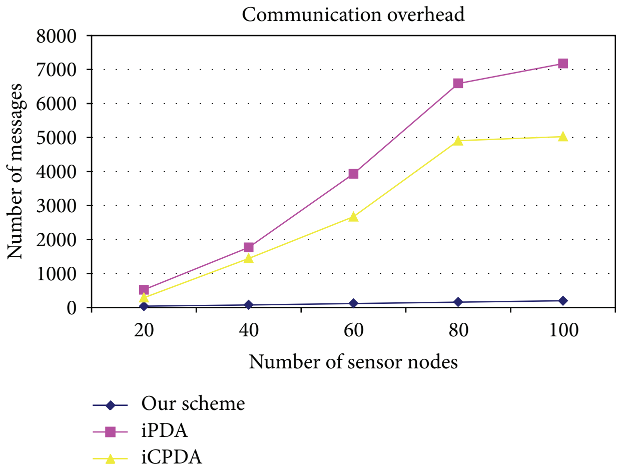


Figure 3: Energy consumption.

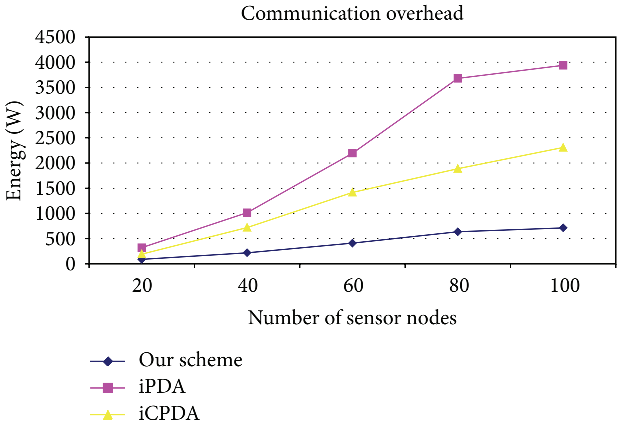


Figure 4: Energy consumption by the iPDA, iCPDA, and our schemes.

4.1. Data Aggregation. Figure 3 shows communication overhead in terms of the number of messages generated in a WSN with respect to varying number of sensor nodes. As expected, the number of messages in the iPDA, iCPDA, and our schemes increases when the number of sensor nodes increases. This is because every sensor node in the WSN is capable of sensing data and when the number of source nodes increases, the number of messages also naturally increases in all of the three schemes. However, our scheme outperforms the iPDA and iCPDA schemes because the existing schemes generate unnecessary messages in the network. The reason is that in our scheme each sensor node can customize its data by itself and it does not need to generate extra messages in the network for data privacy and integrity checking. On the other hand, the iPDA and iCPDA schemes generate six messages and four messages, respectively, for privacy preservation and integrity checking. Due to many messages exchanged among the nodes, the existing schemes cause high data collisions. That is to say, the number of messages generated in the network increases drastically as the number of sensor nodes becomes larger. iPDA and iCPDA schemes consume much energy for successful data transmission, compared with our scheme.

The messages generated in the WSN are finally consumed by the sink node. For this, message transmission and message reception processes are involved. Both processes require significant amount of energy. Figure 4 shows communication overhead in terms of energy dissipation by the iPDA, iCPDA, and our schemes with respect to varying number of sensor nodes in the WSN. As expected, the dissipated energy by all three schemes increases when the number of sensor nodes increases. This is because every message generated in the network requires some amount of energy to reach the sink node. However, the power consumption by our scheme is always lower than that of iPDA and iCPDA schemes. The reason is that the iPDA and iCPDA schemes generate too many unnecessary messages in the WSN while

```
Input: An aggregated WSN and SUM aggregation query
Output: SUM aggregation result
Step 1. Assign node ID and generate signature of the ID
    for all sensor nodes {
        ID = $2^n$;  // where $n = 0, 1, 2, \ldots$
        ID = Signature($2^n$);    $n = n + 1$;}
Step 2. Create customized data from the data of the source nodes
    for all sensor nodes {
        sense ds;
        $a$ = mask(ds, sr); // sr is a unique private seed
        $R'_1$ = genCmpxNum($a$, $bi$);
        $C_j$ = Enc($K_{x,y}$, (ID, $R'_1$));
        transmit($C_j$);}
Step 3. Local integrity checking and applying additive property of complex numbers to get intermediate result of the
customized data
    for every intermediate aggregators {
        for all received customized data {
            Drc($K_{y,x}$, ($C_j$));
            If ($bi$ ! = $b'i$ AND $bi > \delta$) //local integrity checking
             {reject $C_j$; inform_Sink();}
            Else {
                SSig = Superimpose($\mathrm{ID}_1, \ldots, \mathrm{ID}_k$);
                $R'$ = SUM($R'_1, \ldots, R'_k$);
                $C_r$ = Enc($K_{y,x}$, (SSig, $R'$));
                transmit($C_r$);} } }
Step 4. Compute aggregation result at the sink node
    for all receive($C_{\mathrm{rs}}$){Drc($K$, ($C_1$));
        $\mathrm{SUM}_2$ = add ($IR'_1, \ldots, IR'_k$);}
Step 5. Identify contributed sensor nodes, extract actual SUM of the sensors data and check global data integrity at the
sink
    fetch_Nodes_IDs();
    Node_IDs = SuperSig && SSig;
    $\mathrm{SUM}_2$ = disjoin ($\mathrm{SUM}_{2\mathrm{R}}$, $\mathrm{SUM}_{2\mathrm{IM}}$);
    $\mathrm{SUM}_{1\mathrm{R}}$ = Compute (sum of real seeds of contributed nodes);
    SUM = $\mathrm{SUM}_{2\mathrm{R}} - \mathrm{SUM}_{1\mathrm{R}}$;
    If ($\mathrm{SUM}_{2\mathrm{IM}} = B'i$ AND $\mathrm{SUM}_{2\mathrm{IM}} \leq \Delta$)/* global integrity
                                                         checking */
        {return SUM;}
    Else {reject SUM;}
```

Algorithm 1: Algorithm for SUM aggregation with privacy-preservation and integrity checking.

achieving integrity protection and privacy preservation in data aggregation. And also every sensor node becomes active for longer time to communicate all the messages. However, in our scheme, every sensor node can achieve both integrity protection and privacy preservation by comparing the current complex number with the previous one. Hence, the energy consumption of our scheme is reduced by 80% and 60% over the iPDA and iCPDA, respectively.

Table 3 shows the computation overhead of data aggregation. The result shows that iCPDA has the worst performance on the computation overhead for privacy-preserving data aggregation. The reason is that the iCPDA uses a time-consuming encryption method with two seeds to achieve data privacy. On the other hand, the computation cost of our scheme is about two times and 83 times faster than those of the iPDA and iCDPA, respectively. It is shown that our scheme reduces a significant amount of resource (CPU time) usage for achieving private data aggregation. This is because our scheme reduces the number of communication messages by using the additive property of a complex number.

4.2. Data Integrity. Figure 5 shows data propagation delay in terms of average time required by sampled data of sensor nodes to reach to the sink node considering data privacy and integrity checking. During this process, a sensor node in iPDA and iCPDA has to communicate (i.e., transmit and receive) at least six and four messages, respectively. Hence, sensor nodes in both iPDA and iCPDA need more active time to perform all communications than our scheme resulting in very high data propagation delay in the existing work. In this way, dutycycling, which is the percent of time that an entity spends in an active state as a fraction of the total time [41], is also increased in the existing schemes. The iCPDA generates less number of messages than the iPDA but has complex

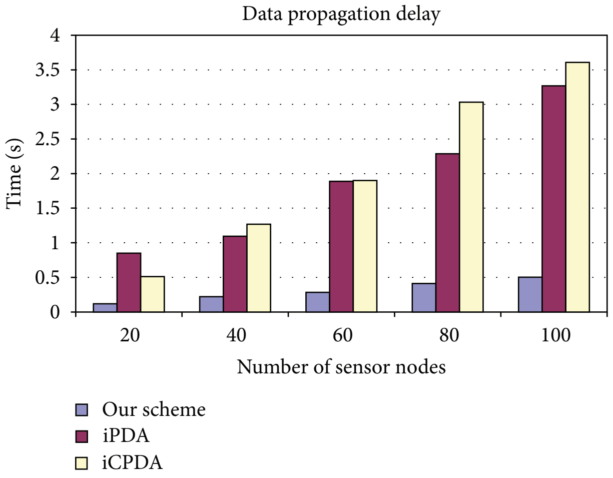


FIGURE 5: Average data transmissions time for iPDA, iCPDA, and our schemes.

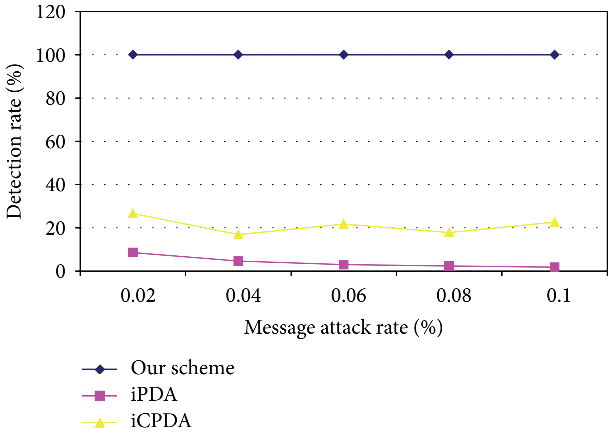


FIGURE 6: Integrity checking.

TABLE 3: Computational overhead for data customization and aggregation.

Protocols	Execution time (in sec)
iPDA	0.005924
iCPDA	0.219325
Our scheme	0.002632

computation for privacy preservation and longer size message than that of the iPDA. Moreover, in iCPDA, the sampled data of sensor nodes is sent to the opposite direction (data is transmitted from the cluster head to the cluster members) of the sink node for privacy preservation process. Therefore, the iCPDA has the worst performance among the three schemes. On the other hand, every sensor node in our scheme sends only one message (the aggregated data) to its parent node because it checks the integrity of the sensed data without the communication of other sensor nodes.

Figure 6 provides the performance of three schemes in terms of the detection ratio of polluted messages for integrity checking. It is shown that our scheme can detect all polluted messages, whereas iPDA and iCPDA can detect less than 30% of polluted messages. The reason is that every node in our scheme checks the integrity of its incoming data received from the lower-level nodes. On the other hand, only the sink node can check the integrity of the aggregated data in iPDA, whereas only the sink node and the cluster heads can perform the integrity checking in iCPDA.

5. Conclusion

In this paper, we proposed an efficient and general scheme in order to aggregate sensitive data protecting data integrity for private data generating environments such as patients' health monitoring application. For maintaining data privacy, our scheme applies the additive property of complex numbers where sampled data are customized and given the form of complex number before transmitting towards the sink node. As a result, it protects the trend of private data of a sensor node from being known by its neighboring nodes including data aggregators in WSNs. Moreover, it is still difficult for an adversary to recover sensitive information even though data are overheard and decrypted. Meanwhile, data integrity is protected by using the imaginary unit of complex-number-form customized data at the cost of just two extra bytes. Through simulation results, we have shown that our scheme is much more efficient in terms of communication and computation overheads, data propagation delay, and integrity checking than the iPDA and iCPDA schemes.

As future work, we will provide more simulation results by designing data integrity and sensitive data-preserving scheme under collusive attacks. Moreover, we will improve our privacy-preserving data aggregation scheme to support MAX and MIN aggregations.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant no. 2013010099). And this research was supported by the Brain Korea 21 PLUS Project, NRF.

References

[1] S. K. Dhurandher, M. S. Obaidat, and M. Gupta, "An acoustic communication based AQUA-GLOMO simulator for underwater networks," *Human-Centric Computing and Information Sciences*, vol. 2, article 3, 2012.

[2] H. Karl and A. Willig, "A short survey of wireless sensor networks," Tech. Rep. TKN-03-018, 2003.

[3] K. Romer, "Programming paradigms and middleware for sensor networks," in *Proceedings of the GI/ITG Workshop on Sensor Networks*, pp. 49–54, Karlsruhe, Germany, 2004.

[4] Q. Liu and D. Oh, "Performance evaluation of multi-hop communication based on a mobile multi-robot system in a subterranean laneway," *Journal of Information Processing Systems*, vol. 8, no. 3, pp. 471–482, 2012.

[5] G. Carvalho, I. Woungang, A. Anpalagan, and S. Dhurandher, "Energy-efficient radio resource management scheme for heterogeneous wireless networks: a queueing theory perspective," *Journal of Convergence*, vol. 3, no. 4, pp. 15–22, 2012.

[6] X. Li, N. Mitton, A. Nayak, and I. Stojmenovic, "Achieving load awareness in position-based wireless Ad Hoc routing," *Journal of Convergence*, vol. 3, no. 3, pp. 17–22, 2012.

[7] M. Yoon, Y. K. Kim, and J. W. Chang, "An energy-efficient routing protocol using message success rate in wireless sensor networks," *Journal of Convergence*, vol. 4, no. 1, pp. 15–22, 2013.

[8] B. Singh and D. Lobiyal, "A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks," *Human-Centric Computing and Information Sciences*, vol. 2, article 13, 2012.

[9] R. Sumathi and M. G. Srinivas, "A survey of QoS based routing protocols for wireless sensor networks," *Journal of Information Processing Systems*, vol. 8, no. 4, pp. 589–602, 2012.

[10] M. S. Obaidat, S. K. Dhurandher, and K. Diwakar, "CASPER: congestion aware selection of path with efficient routing in multimedia networks," *Journal of Information Processing Systems*, vol. 7, no. 2, pp. 241–260, 2011.

[11] J. Considine, F. Li, G. Kollios, and J. Byers, "Approximate aggregation techniques for sensor databases," in *Proceedings of the 20th International Conference on Data Engineering (ICDE '04)*, pp. 449–460, April 2004.

[12] M. Samuel, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a tiny aggregation service for ad hoc sensor networks," in *Proceedings of the Symposium on Operating Systems Design and Implementation (OSDI '02)*, 2002.

[13] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on Database Systems*, vol. 30, no. 1, pp. 122–173, 2005.

[14] R. Bista, Y.-K. Kim, and J.-W. Chang, "A new approach for energy-balanced data aggregation in wireless sensor networks," in *Proceedings of the 9th IEEE International Conference on Computer and Information Technology (CIT '09)*, vol. 2, pp. 9–15, October 2009.

[15] F. Tseng, L. Chou, and H. Chao, "A survey of black hole attacks in wireless mobile ad hoc networks," *Human-Centric Computing and Information Sciences*, vol. 1, article 4, 2011.

[16] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "PDA: privacy-preserving data aggregation in wireless sensor networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 2045–2053, May 2007.

[17] T. Feng, C. Wang, W. Zhang, and L. Ruan, "Confidentiality protection for distributed sensor data aggregation," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 475–483, April 2008.

[18] M. Conti, L. Zhang, S. Roy, R. Di Pietro, S. Jajodia, and L. V. Mancini, "Privacy-preserving robust data aggregation in wireless sensor networks," *Security and Communication Networks*, vol. 2, no. 2, pp. 195–213, 2009.

[19] W. He, H. Nguyen, X. Liu, K. Nahrstedt, and T. Abdelzaher, "iPDA: an integrity-protecting private data aggregation scheme for wireless sensor networks," in *Proceedings of the IEEE Military Communications Conference (MILCOM '08)*, pp. 1–7, November 2008.

[20] E. Mlaih and S. A. Aly, "Secure hop-by-hop aggregation of end-to-end concealed data in wireless sensor networks," in *Proceedings of the IEEE INFOCOM Workshops*, pp. 1–6, April 2008.

[21] W. He, X. Liu, H. Nguyen, and K. Nahrstedt, "A cluster-based protocol to enforce integrity and preserve privacy in data aggregation," in *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops*, pp. 14–19, 2009.

[22] E.-O. Blaß and M. Zitterbart, "An efficient key establishment scheme for secure aggregating sensor networks," in *Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS '06)*, pp. 303–310, March 2006.

[23] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 439–450, Dallas, Tex, USA, May 2000.

[24] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques," in *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM '03)*, pp. 99–106, Melbourne, Australia, November 2003.

[25] M. Conti, L. Zhang, S. Roy, R. Di Pietro, S. Jajodia, and L. V. Mancini, "Privacy-preserving robust data aggregation in wireless sensor networks," *Security and Communication Networks*, vol. 2, no. 2, pp. 195–213, 2009.

[26] J. Yao and G. Wen, "Protecting classification privacy data aggregation in wireless sensor networks," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, pp. 1–5, Dalian, China, October 2008.

[27] W. Zhang, C. Wang, and T. Feng, "GP^{2}S: generic privacy-preservation solutions for approximate aggregation of sensor data," in *Proceedings of the 6th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom '08)*, pp. 179–184, Hong Kong, China, March 2008.

[28] M. Yoon, K. Yong-Ki, and J. Chang, "A new data aggregation scheme to support energy efficiency and privacy preservation for wireless sensor networks," *International Journal of Security & Its Applications*, vol. 7, no. 1, pp. 129–142, 2013.

[29] Y. K. Kim, H. Lee, M. Yoon, and J. W. Chang, "Hilbert-curve based data aggregation scheme to enforce data privacy and data integrity for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 217876, 14 pages, 2013.

[30] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 457–465, Phoenix, AZ, USA, April 2008.

[31] 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)*, vol. 5, pp. 3044–3049, Seoul, Korea, May 2005.

[32] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous Systems -Networking and Services (MobiQuitous '05)*, pp. 109–117, San Diego, Calif, USA, July 2005.

[33] G. Taban and D. Gligor, "Privacy-preserving integrity-assured data aggregation in sensor networks," in *Proceedings of the IEEE International Conference on Privacy, Security, Risk, and Trust (PASSAT '09)*, pp. 168–175, Vancouver, Canada, August 2009.

[34] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous Systems -Networking and Services (MobiQuitous '05)*, pp. 109–117, July 2005.

[35] E.-O. Blaß and M. Zitterbart, "An efficient key establishment scheme for secure aggregating sensor networks," in *Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS '06)*, pp. 303–310, March 2007.

[36] J. Zobel, A. Moffat, and K. Ramamohanarao, "Inverted files versus signature files for text indexing," *ACM Transactions on Database Systems*, vol. 23, no. 4, pp. 453–490, 1998.

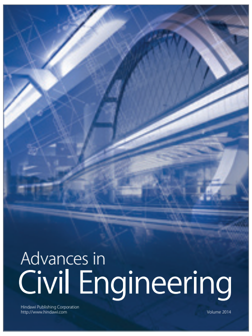
[37] R. Bista and J. W. Chang, "Energy efficient data aggregation for wireless sensor networks," in *Sustainable Wireless Sensor Networks*, 2010.

[38] P. Levis, N. Lee, M. Welsh, and D. Cullar, "TOSSIM: accurate and scalable simulation of entire TinyOS applications," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*, pp. 126–137, 2003.

[39] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K. Pister, "System architecture directions for networked sensors," in *Proceedings of the 9th Internatinal Conference Architectural Support for Programming Languages and Operating Systems (ASPLOS '00)*, pp. 93–104, November 2000.

[40] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM '00)*, pp. 56–67, August 2000.

[41] http://en.wikipedia.org/wiki/Duty_cycle.



Hindawi

Submit your manuscripts at
http://www.hindawi.com