Ann. Telecommun. (2016) 71:465–475
DOI 10.1007/s12243-016-0498-7

A differential privacy protection scheme for sensitive big data in body sensor networks

Chi Lin[1,2] · Pengyu Wang[1,2] · Houbing Song[3] · Yanhong Zhou[1,2] · Qing Liu[1,2] · Guowei Wu[1,2]

Received: 5 September 2015 / Accepted: 26 February 2016 / Published online: 9 March 2016

Abstract As a special kind of application of wireless sensor networks, body sensor networks (BSNs) have broad application perspectives in health caring. Big data acquired from BSNs usually contain sensitive information, such as physical condition, location information, and so on, which is compulsory to be appropriately protected. However, previous methods overlooked the privacy protection issue, leading to privacy violation. In this paper, a differential privacy protection scheme for sensitive big data in BSNs is proposed. A tree structure is constructed to reduce errors and provide long range queries. Haar Wavelet transformation method is applied to convert histogram into a complete binary tree. At last, to verify the advantages of our scheme, several experiments are conducted to show the outperformed results. Experimental results demonstrate that the tree structure greatly reduces the calculation overheads which preserves differential privacy for users.

Keywords Sensitive information · Body sensor networks · Differential privacy protection

1 Introduction

As a special application of wireless sensor networks (WSNs) [1], body sensor networks (BSNs) [12, 32] are deployed on the surface of bodies for periodically monitoring physical conditions [24]. In some cases, especially in emergency or health care, security and privacy properties are extremely important [25]. Because a slight leakage of sensitive data may cause unpredictable damages. Therefore, extensive studies on privacy preservation have been carried out, which is one of the most critical research topics in BSNs [16].

Usually, data collected, aggregated, and transmitted in BSNs contain personal and sensitive private information [11], which directly or indirectly reveals the condition of a person [26]. If the data cannot be properly preserved, once exposed to the public, the privacy will be destroyed. Therefore, protecting the privacy of sensitive data is of great importance [2, 20].

In general, traditional methods for protecting privacy and security of big data in BSNs fall into three categories [6–8, 18, 21, 23, 27]: (1) anonymous techniques, (2) privacy protection rules, and (3) collaborative filtering.

However, the above-mentioned privacy protection schemes are still suffering from several common problems:

1. Anonymous technologies [23, 27] simply hid or replaced the information such as identity or location. They overlooked the fact that an attacker can identify a certain user based on his background knowledge

✉ Chi Lin
chilin@mail.dlut.edu.cn

Pengyu Wang
wgwdut@dlut.edu.cn

Houbing Song
h.song@ieee.org

[1] School of Software, Dalian University of Technology, Dalian, China

[2] Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian, China

[3] Department of Electrical and Computer Engineering, West Virginia University, Montgomery, WV 25136, USA

although an objective user's identity information is concealed or deleted. Obviously, anonymous technique does not protect the accuracy or availability of data.

2. Privacy protection rules [6, 10, 21] mainly aimed at finding out the underlying relationship among data. However, this method can only cope with a fixed mode of attack.
3. Collaborative filtering schemes [7, 8, 14] loosed degrees of the data set and calculated errors based on similarity. They assumed that the attacker cannot grasp full background knowledge. Such limitations seriously constraints the widespread application based on this scheme.

Researching the mechanism of privacy protection in BSNs provides fundamental contributions in developing network security technology. With the growing demands for wearable devices, privacy protection of big data has become a major concern. As the big data era comes, personal information leakage happens more and more frequently, such as the famous cookies storm happened several years ago, which caused 360 users' information exposed. Therefore, privacy protection of BSNs should be paid close attention to [16].

Our motivation is to design a scheme to protect the privacy of BSNs based on differential privacy technology [9, 30]. In our scheme, we change the structure of the data set to reduce the sensitivity, and then add noise to the deformation data set, and finally, get the published data set [36].

The contribution of this paper can be summarized as follows:

1. To the best knowledge of the authors, this is probably the first time that differential privacy combining with Haar Wavelet technique is applied to protect the privacy in BSNs. We develop a tree-based structure to analyze the data so as to reduce the error and provide long range queries.
2. In order to add noise conveniently, we use Haar Wavelet transform method [30] to convert histogram into a complete binary tree.
3. To verify the advantages of our scheme, several experiments are conducted to show the outperformed results. Experimental results reveal that our scheme greatly reduces the overhead of calculations.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the state of the art of the privacy preservations for BSNs. Section 3 introduces related preliminaries of this paper. In Section 4, a differential privacy protection scheme is proposed, which aims at protecting the sensitive information of the BSNs. Experiments are conducted in Section 5. Finally, we conclude this paper and suggest our future work in Section 6.

2 Literature review

With the continuous interest in BSNs for wearable device, the growing emergence of new techniques inspired great efforts on the research of the privacy protections in BSNs [2, 15]. In literature, particular attentions are paid for protecting sensitive data in BSNs. In general, the approaches of preserving the privacy of big data fall into three categories.

Barua et al. [5] proposed packet scheduling schemes for real-time transmission in wireless body area networks (WBANs) with proper security and privacy. Real-time and non-real-time traffic are classified to minimize the waiting time of the eHealth application's data traffic. Liu et al. [19] presented a remote anonymous authentication protocol to enable client terminals/application to access WBANs services securely. Trcek et al. [28] proposed address privacy for the Internet of things technology by focusing on the most primitive members, bare sensors and RFIDs. A strategy of incrementally adjusting existing protocols is adopted. Antonescu et al. [4] provided a comprehensive review of the challenges and emerging technologies for WBANs.

To provide availability, integrity, and confidentiality for data, privacy protections in BSNs mainly concentrated on cryptography technology. He et al. [13] presented the design, implementation, and evaluation of a secure network admission and transmission subsystem based on a polynomial-based authentication scheme. The procedures to establish keys for each biosensor in this subsystem are communication efficient and energy efficient. Ali et al. [3] demonstrated a scheme which is able to construct shared keys with near-perfect agreement for the secret key generation, avoiding the cost on reconciliation. Zhao et al. [34] proposed an open research issue that should be solved in the future key negotiation protocols. They explore and classify these solutions, and evaluate their performance by analyzing their merits and demerits. Li et al. [17] proposed group device pairing (GDP), a user-aided multi-party authenticated key agreement protocol. Based on GDP, a group of sensor devices which have no pre-shared secrets establish initial trust by generating various shared secret keys out of an unauthenticated channel. From the above several methods, we know that cryptography technology is better than privacy-protected methods in WBANs to be applied in BSNs. However, to reduce energy consumption, most of them just use cryptography technology without considering the data sets. Moreover, the encrypted data are prone to attention of the attacker and be the target of interception and attack. With the appearance of new attack methods, the scope of applicability of encryption technology will be smaller. Therefore, protecting security and privacy of the sensitive data in BSNs is still a challenging problem.

In WSNs, another research proposed a kind of method that build a particular transport protocol between the communications of sensor network device to protect the data sets. Yao et al. [33] designed a regulator packing real data session into an independent transmission model at transmission layer. All valid data packets are equably sent at the frequency defined by the regulator and at the same length to clutter the inherent pace of valid data transmission and other parameters. They also proposed a strategy PAS to minimize the overhead while preventing attackers from locating the patients. Lu et al. [22] proposed PSSS, Physiological Signals based Secret Sharing scheme, aiming at deploying identical secrets automatically among the nodes of BSNs, which is considered as an important add-on security mechanism in BSNs. By utilizing biometric characteristics of physiological signal, PSSS can be efficient, reliable, and free of third-party authentication and pre-distribution. Yan et al. [31] proposed a novel In-network AES Equivalent (IAE) mechanism to protect the security/privacy and maintain good energy efficiency for WBASNs at the same time. IAE achieves this goal by outsourcing part of the energy-consuming cryptographic operation to other deliberately selected peer sensor nodes, so as to balance the energy consumption of the entire network. In [35], Zhou et al. mainly focused on the goals and tactics of privacy-preserving data aggregation in cloud-assisted wireless wearable communications. With respect to the unique security and privacy requirements and the efficiency consideration for resource-constrained wearable devices, they identified the inappropriateness of secure multiparty computation and fully homomorphic encryption. Venkatasubramanian et al. [29] presented physiological value-based security (PVS), a usable and efficient way of securing inter sensor communication schemes for BSNs. The PVS scheme distributes the key used for securing a particular message along with the message itself, by hiding it using physiological values. In this way, it not only eliminates the need for an explicit key distribution but also reduces the number of keys required at each node to meet all its secure communication requirements.

Although all the above privacy protected schemes provide solutions to improve the security and privacy of BSNs, most of them just protect the data from outside and ignore data encryption and transport protocol on privacy; once there is a new attack method that appears, the limitations of these methods will be revealed and the goal that protect the data permanently will not be achieved [2]. Imagine, if there is some VIP data information leaked, the result will be terrible. In this paper, a privacy protection scheme based on differential privacy is developed to avoid the situation mentioned above.

The major differences between this work and the aforementioned schemes are that differential privacy protection is a new definition of privacy protection. Differential privacy protection considers the background knowledge of the attacker to protect user's privacy information of a protection mechanism. The general idea is inserting or deleting one user's data will not affect the availability. In that case, even if the attacker masters more information, they cannot figure out which data belongs to the target user.

3 Preliminaries

In this section, related preliminaries are introduced.

3.1 Differential privacy

The main idea of the differential privacy protection is demonstrated as follows.

If a data set D includes a message of Bob, then various operations $M(D)$, such as average, count, and so on, are performed to obtain statistical results. Suppose that after Bob's information is deleted from D, which changes D and into D', performing $M(D')$ or $M(D)$ yields almost identical results, it is considered that Bob's information in the data set D under the operation $M(D)$ is safe. Because either missing his information or not does not affect the output.

Then, we introduce the definition of proximate data.

Suppose there is a finite field Z and a data set D_1, all elements in D_1 are made up of the elements in Z. The number of elements of D_1 is denoted as n and the number of attributes of D_1 is denoted as d. We define f as a query of the data set D_1 and F represents the collections of f. For privacy protection concerns, we use algorithm A to encrypt the results generated by F. Suppose there is a data set D_2 which has the same structure as D_1 (i.e., $A(D_1)$). $D_1 \Delta D_2$ represents the difference between the two data sets, and $|D_1 \Delta D_2|$ represents the number of elements in $D_1 \Delta D_2$. If $|D_1 \Delta D_2| = 1$, D_1 and D_2 are considered as proximate data.

Definition 1 Differential privacy. Given two similar data sets D_1 and D_2, and a privacy protection algorithm A. If the results of $A(D_1)$ or $A(D_2)$ is O, P_r represents the probability of privacy loss, and O satisfy the following inequation:

$$P_r[A(D_1) = O] \leq e^z * P_r(A(D_2) = O), \tag{1}$$

then algorithm A is regarded as satisfying *ε-difference privacy*.

Here, ε is called differential privacy budget, which is defined below.

Definition 2 Differential privacy budget ε. Given two data set, D_1 and D_2 , the ratio of two results operated by algorithm A is denoted as ε (see Eq. 2).

$$\varepsilon = \frac{A(D_1)}{A(D_2)} \tag{2}$$

Figure 1 shows the risk of privacy leakage. We note that two similar data sets D_1 and D_2 meet *ε-difference privacy* when the risk is very small.

When the value of ε approaches to 1, the availability of D_1 and D_2, which are processed by the algorithm A, will remain high, whereas the level of privacy protection is low. Lower value of ε indicates higher level of privacy protection, and furthermore, the processed data set vary widely. When $\varepsilon = 0$, maximum privacy protection is achieved, however, at this point, data become unavailable. Therefore, ε should not be as small as possible. We should appropriately balance the data availability and privacy according to actual situation.

Definition 3 Sensitivity of differential privacy Δf : A usual method of achieving difference privacy is adding noise. Here, we consider how to trade off the amount of noise to be added into the original data. When adding too much noise, the data will be far from the original value, destroying the accuracy. On the contrary, when little noise is added, it will not be able to protect data properly. As an important indicator of adding noise, the value of differential privacy sensitivity Δf is denoted as the change of data set we get after deleting any records of the data set, which is calculated by Eq. 3.

$$Deltaf = \max_{D_1, D_2} ||f(D_1) - f(D_2)|| \tag{3}$$

3.2 Adding noise

First of all, we demonstrate the importance of adding noise by a simple example.

Four patients' status are listed in the data set of hospital medical records as shown in Table 1. It shows whether a patient catches a cold. In the table, 1 denotes catching a cold and 0 denotes health.

If the data set provides query services, we define $f(i) = num(i)$ as the number of the patients among i people. For example, $f(1) = 1$, $f(3) = 2$, $f(4) = 3$. We also assume that it is impossible to find out the prevalence of a certain patient. However, these two requirements cannot achieve protecting privacy. For example, one attacker wants to get Jack's prevalence, therefore, he can know Jack is sick by calculating the value of $f(4) - f(3) = 1$.

Data set D1
Dataset D2

Fig. 1 Neighbor dataset

Table 1 Diagnostic results in hospital medical records

Name	Diagnostic results
Tom	1
Bob	0
Mary	1
Jack	1

To avoid privacy leakage, we additionally put noise into function f. We denote $f'(i) = num(i) + noise$. Afterward, results obtained by $f'(i)$ can be $\{2, 2, 2, 3\}$ or $\{0, 1, 2, 2\}$, etc., which thus protect the privacy of all patients within the group. Moreover, after adding the noise, the diagnosis of Jack became unavailable.

Adding noise is a main technology of differential privacy to achieve privacy protection. There are two common methods to add noise: (1) Laplace mechanism, and (2) Indexing mechanism.

Laplace mechanism After adding random noise, which obeys Laplace distribution to the results generated from the data set, the effect of privacy protection will be achieved. We need to add noise function $Lap(b)$ with probability density $p(x) = exp(-|x|/b)/2b$, $Lap(b) = exp(-x/b)$, here, b satisfies $b = \Delta f/\varepsilon$ and algorithm $A(D) = f(D) + Y$. The result of the query $f(D)$ is $Y \sim Lap(\Delta f/\varepsilon)$.

Indexing mechanism In many cases, Laplace cannot satisfy the query of object, because it only queries the value. When we need to query a choice or a solution, we usually choose the indexing mechanism. For example, we use the indexing mechanism to quantify the degree of how much we agree with a scheme.

4 Our scheme

In this section, we illustrate our scheme for preserving privacy for sensitive big data in BSNs by utilizing differential privacy technology in details.

4.1 Tree structure-based scheme

In our scheme, we use a non-interactive method to realize differential privacy protection. This approach helps to reduce sensitivity and makes it difficult to destroy data structures. Specific methods are shown in Algorithm 1.

Algorithm 1 Privacy protection strategy

```
Input: Dataset $D$
Output: Dataset $D''$
Deformation new data set $D'$
Add noise into $D'$ and convert it into $D''$
Return: Dataset $D''$
```

Our scheme is to perform a treatment to the data set. Usually, a structure transformation is conducted through the structure deformation to decrease the sensitivity. Then, the noise is added to the data set for obtaining a release of the data sets after deformation. We build a histogram for recording the deformation after adding noise. Therefore, it can reduce errors and can provide remote query.

The key to our scheme is selecting an appropriate structure to analyze the data. Through the analysis and comparison, we adopt a tree hierarchy. Each group of histogram serves as a leaf node of the tree, and we number the node id according to the order of histogram from left to right.

Then, a complete tree is structured. Based on the concept discussed previously, we know that apart from budget ε, sensitivity Δf is also another main factor, which is created by deleting or adding data into the data set; this also causes the key error for differences privacy protection. The same histogram has the same number of leaf nodes in the tree. If a different node degree is used, it will generate different height of trees; for example, a tree with eight leaves will produce the complete binary tree with a height of four and a ternary tree with three. For the original data set D, deleting data to obtain D' will only affect a leaf node. Thus, the sensitivity Δf equals to the height of the tree. In this example, based on the sensitivity of a binary tree $\Delta f = 4$ and ternary tree $\Delta f = 3$, the larger the value of Δf is, the greater the effects will be generated.

Then, we illustrate how to construct a tree structure and how to set corresponding values based on histogram.

Firstly, for a complete binary tree, Haar Wavelet transform is used for grouping histogram. The Haar Wavelet function is discontinuous, similar to the step function, which is defined as:

$$\varphi(x) = \begin{cases} 1 & 0 \le x \le 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & Others \end{cases} \tag{4}$$

The scale function is:

$$\phi(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & Others \end{cases} \tag{5}$$

For a complete binary tree, the left node of the range is considered as $x(0, 1/2)$, while the right node of the range is $x(1/2, 1)$. Through the Haar Wavelet transformation, it converts histogram into a complete binary tree.

Table 2 Frequency for each group

Age group	The number of people
0 ∼ 10	2
10 ∼ 20	4
20 ∼ 30	2
30 ∼ 40	6

Suppose there are n groups of histograms. We use the Haar Wavelet transformation method to convert the histogram into a complete binary tree and optimize the structure. Each node of the complete binary tree will generate n lines with differential privacy groups by adding Laplace operation. Data released from n groups are not only available but also can stop an attacker from getting the desired information.

For a group of data set, each id number in the investigation of medical corresponds to the heart rate. First, histogram data are optimized into several groups, each group matches the corresponding age range. Each heartbeat data point corresponds to the appropriate range and the frequency range will plus 1. Moreover, each group represents the frequency of the band's original data in the resulting histogram.

In order to describe the algorithm better, we assume that $n = 8$. The original data set $D = \{2, 4, 2, 6\}$ is shown in Table 2.

For convenience, the first step takes $\varepsilon = 0.5$ to indicate the availability. It can be seen from Fig. 2 that the height of the binary tree is 3 and the sensitivity is $\varDelta f = 3$.

Next, we introduce the process of our algorithm to get the tree structure.

Firstly, the established Knode structure is used to store the original data set. For example, when the original data set has 3600 pieces of data, it initiates an information array to store the original data set. Then we find the maximum and minimum values. We calculate the difference between the

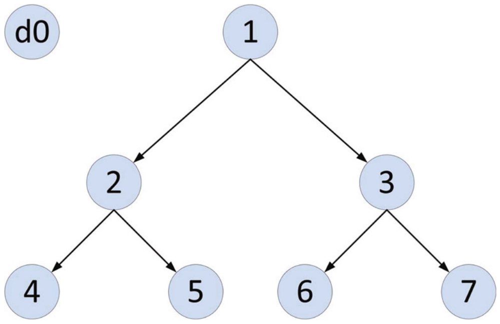

Fig. 2 A complete binary tree for Haar Wavelet

Fig. 3 Shimmer body sensor networks

two values divided by the number of groups. Secondly, the width of the group is divided by the function which creates a node to convert the original data in each group into a histogram. Each group stores the frequency. The value of the node is the original data for each group in the histogram. It is referred to as the distribution of data privacy protection, which will be used later in this work.

In the initialization of all non-leaf nodes of the tree: d_0, d_1, d_2, and d_3, are created, where d_0 stores the average value of frequency of leaf nodes in the complete binary tree and d_1, d_2, and d_3 are the wavelet coefficients. The id of the d_i wavelet coefficients is calculated as follows: The average value of all leaf nodes in the left subtree of d_i is denoted as a, and the average value of all the leaf nodes in the right subtree of d_i is denoted as b, making the d_i values as follow:

Table 3 Original data

ID	Value	ID	Value
180	70.111115	197	68.079369
198	87.854698	199	71.616608
200	77.471306	201	69.299144
202	83.119659	190	64.551892
191	78.184372	192	90.095238
193	87.451767	194	81.742371
195	62.956043	196	77.472527
203	81.307693	204	85.152626
205	66.409035	206	80.169716
181	65.780220	182	60.064713
183	62.440781	184	92.256409

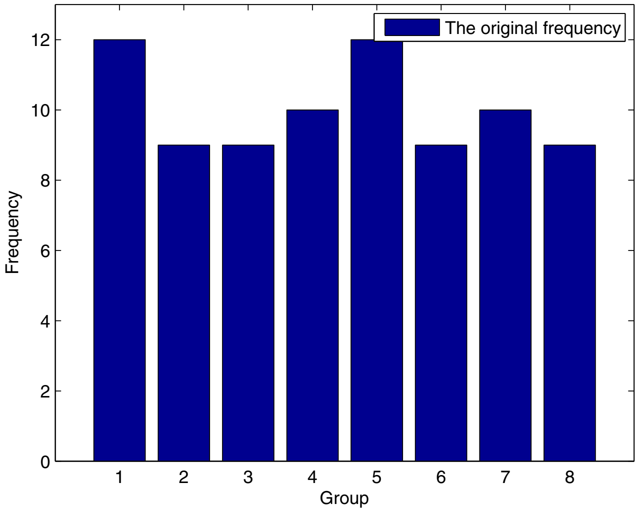

Fig. 4 Original frequency

$$d_i = (a - b)/2 \tag{6}$$

In our example, we have $d_2 = (2 - 4)/2 = -1$. d_0 is the average value of all the leaf nodes, where ($d_0 = 3.5$). Then we add noise to the binary tree for each wavelet coefficients. Here, we use Laplace equation as:

$$Lap(b) = exp(-|x|/b) \tag{7}$$

From $b = \Delta f/\varepsilon$, we note the mainly influence of noise mainly relies on values of Δf and ε. ε is a constant, and Δf is a difference value between original data set and newly generated data set, which closely relates to the height of the tree. Then noise added by wavelet can be calculated as:

$$Lap((1 + \log_2 n)/(\varepsilon * W_{Haar}(d_i))), \tag{8}$$

Table 4 Original frequency

ID	Frequency	Heartbeat (beats/min)
8	12.000000	62.546703
9	9.000000	67.539986
10	9.000000	72.533272
11	10.000000	77.526558
12	12.000000	82.519844
13	9.000000	87.513123
14	10.000000	92.506409
15	9.000000	97.499695

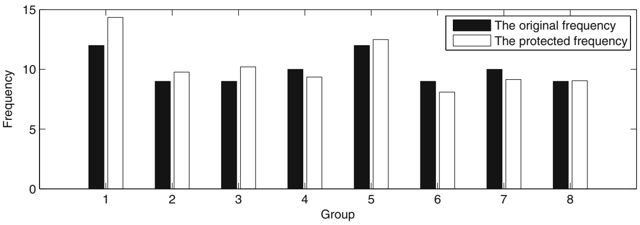

Fig. 5 Comparison of frequency before and after treatment

here, $W_{Haar}(d_i) = 2^{h-i+1}$, h is denoted as the height of the complete binary tree. i denotes the layers which d_i locates in the binary tree. Then Laplace transformation is executed to assign noise or each coefficient d_i. In our example, the amount of noise added for d_2 is $Lap(1 + \log_2 4/4\varepsilon)$.

After receiving all wavelet coefficients of the complete binary tree, we calculate the value recorded in each leaf node. We denote the leaf nodes as $b_1, b_2 ... b_n$. We have:

$$b_i = c_0 + \Sigma_{i=1}^{h} f_i * c_i \tag{9}$$

Here, c_i is the value of wavelet coefficient and, the values of f_i are associated with the left and right subtrees of b_i. Starting from the first layer of the complete binary tree with height h based on the above formula, if the leaf node b_i belongs to the left subtree of c_i, then $f_i = 1$; if it is below the right subtree, then $f_i = -1$; and if it is not a c_i subtree, we will simply ignored it without processing.

So far, we introduce how to use the Haar Wavelet coefficients to add noise to the complete binary tree leaf, which changes the structure of the histogram to obtain the effect of differential privacy. We know that the smaller the value of differential privacy budget ε is, the greater the noise will generate, although the results has been better protected, the availability of data is significantly reduced, so the difference in budget ε is an important parameter.

Table 5 Frequency of added noise

ID (Heartbeat)	Frequency
8 (62.546703)	14.348530
9 (67.539986)	9.770548
10 (72.533272)	10.210751
11 (77.526558)	9.355263
12 (82.519844)	12.500614
13 (87.513123)	8.097567
14 (92.506409)	9.150123
15 (97.499695)	9.000000

Table 6 Frequency contrast

ID	Original frequency	Handled frequency
8	12.000000	14.348530
9	9.000000	9.770548
10	9.000000	10.210751
11	10.000000	9.355263
12	12.000000	12.500614
13	9.000000	8.097567
14	10.000000	9.150123
15	9.000000	9.050000

4.2 Privacy protection algorithm

The above operation can be used in each group to satisfy the original histogram of differential privacy. The main algorithm is described in Algorithm 4.2.

Algorithm 2 Privacy protection algorithm

```
Convert original data into $d_i$. Each $d_i$ stores the corresponding frequency.
$d_0$ stores the average value of all leaves of the complete binary tree.
Haar Wavelet method is used to assign values for each $d_i$.
Add Laplace noise for non-leaf nodes $d_i$ based on the binary tree structure.
After getting wavelet coefficients, find the path of each root leaf node $d_0$.
Save frequency for each node into files by inducing wavelet coefficients of each leaf node.
```

In addition to the differential privacy budget ε, the sensitivity Δf is another main parameter of differential privacy. Sensitivity Δf is the maximum difference when calculating results after deleting records in a data set. In the tree structure, the change of leaf node will affect the maximum value of other nodes which depend on the height of the tree. When the number of leaves is fixed, the height of the tree will be

Table 7 Deleted information

ID	Heartbeat (beats/min)	ID	Heartbeat (beats/min)
67	70.918190	68	83.521370
69	87.653236	70	93.511592
71	89.065933	72	79.401711
73	68.216118	74	89.755798
75	78.742371	76	78.322342
77	97.974358	78	89.783882

Table 8 Frequency of deleted data

ID	Frequency	Heartbeat (beats/min)
8	11.000000	62.546703
9	9.000000	67.539986
10	9.000000	72.533272
11	10.000000	77.526558
12	12.000000	82.519844
13	9.000000	87.513123
14	10.000000	92.506409
15	9.000000	97.499695

Table 9 Deleted information

ID (Heartbeat)	Frequency
8 (62.546703)	13.219845
9 (67.539986)	9.472412
10 (72.533272)	9.693085
11 (77.526558)	8.945652
12 (82.519844)	13.054348
13 (87.513123)	8.306915
14 (92.506409)	9.527588
15 (97.499695)	9.000000

determined by the degree of tree m. When $m = 2$, a complete binary tree will be established. When $m = 3$, a ternary tree will be created.

5 Experiment results

In this section, experiments are conducted to verify the advantages of the proposed schemes.

5.1 Test of differential privacy

In this paper, the data of human's heartbeat were collected by a number of Shimmer wearable sensors (http://http://www.shimmersensing.com/), which is shown as Fig. 3. Several experimental data are listed in Table 3.

The experimental environment was Windows 7 operating system and the compiler was codeblocks, data were stored as text files (.TXT), and MATLAB simulation was used. Since our scheme is based on the histogram, we collected the data and then convert them into meaningful histograms as shown in Fig. 4. In our experiment, data are collected from different ages of people such as 60–70, 70–80, and so on. We divided data into groups which are the basis for creating histogram. For ease of analysis, we selected data set of 80 people for our experiments.

The Haar Wavelet transformation method is used in the histogram based on a complete binary conversion. The results are listed in Table 4. Here, the first column indicates the node number, the second column is the frequency, and the third column is the middle value for each range.

Figure 5 is depicted to show comparison of frequency between the original data and the processed data. Table 5 shows the encrypted data node information. As the former seven nodes store the wavelet coefficients, we only list the data to be protected numbered from 8 to 15. Table 6 lists the comparison of values before and after differential privacy protection.

As shown in the Fig. 5, blue bars represent the raw data, red bars stand for the data obtained by performing the differential privacy protection. We note that values move up and down and the error is not big. Therefore, when the attacker grasp full background knowledge, we cannot still identify the objective user because of the noise adding technology. Hence, our scheme can provide sufficient privacy protections and the personal heartbeat data can be successfully protected.

Next, we test whether this method satisfies the differential privacy protection requirements. If two sets are approximate data and the results of the statistical inquiry are almost identical. In this way, it can meet the differential privacy protection.

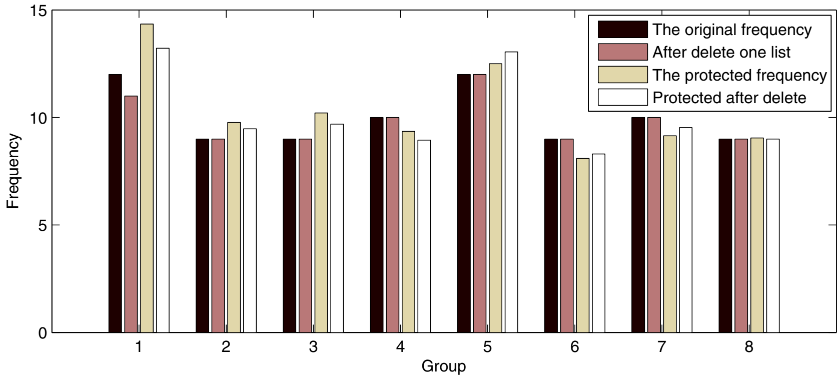

Fig. 6 Deleted difference comparison of data generated

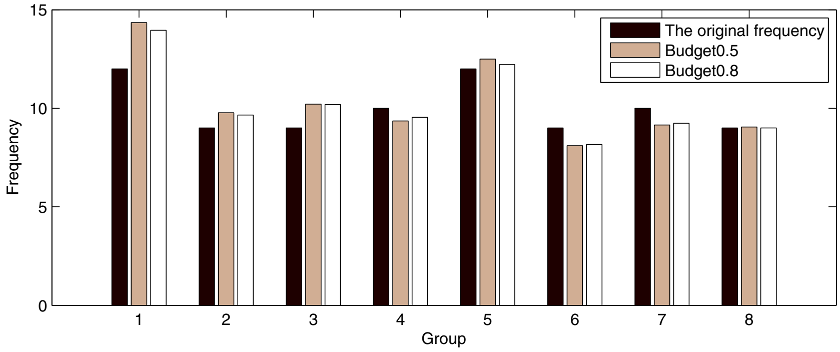

Fig. 7 Comparison of different values from Å

Among the 80 pieces of data, in order to facilitate our notion, we chose to delete the last message, as shown in Table 7. We deleted the data numbered 79.

After deleting data, we perform privacy differential protection scheme. We compare the results before and after deleting data.

Table 8 shows the data after deleting individual nodes as well as the frequency corresponding to the number of heartbeat values.

Figure 6 below each diagram shows the frequency of each node after deleting a record and adding the noise (see Table 9). We observe that our scheme is able to achieve differential privacy protection.

From the histogram, we note that, when differential privacy protection is not applied, deleting one data in the data set will only influence the frequency of one group. The attacker will immediately find out which user's data is deleted. On the contrary, when differential privacy protection is used, the influences of deleting one data are migrated. Data generated by our algorithm still have high availability. Moreover, data frequencies of all groups have been changed in different extents. Therefore, the attacker cannot distinguish the range of the deleted data, which achieves privacy protection. Therefore, we can conclude that the results obtained by our algorithm can protection the privacy of the data while guaranteeing data availability (Table 9).

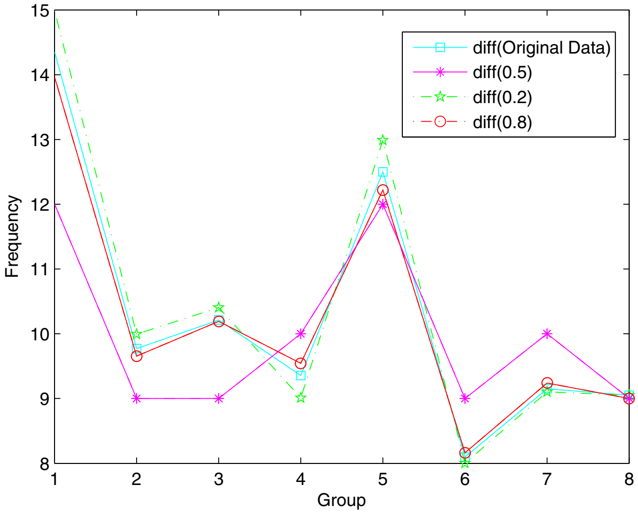

Fig. 8 Differential budget for privacy-treated node frequency

5.2 Impact of budget for noise

Next, we testify whether data encryption will cause an effect when the budget ε changes. As previously mentioned, the value of ε will influence the effect of differential privacy. When the value of ε is close to 1, the data availability will be higher. When ε is close to 0, the protective effect is the best, but the data will become unavailable. In the previous experiments, the value of ε was set to 0.5. Now we make $\varepsilon = 0.8$ and compare the results of these two experiments. As shown in Fig. 7, the blue bar represents the original data, the green bar indicates $\varepsilon = 0.5$, and red bar stands for $\varepsilon = 0.8$. Compared to the green line, the red line is closer to the blue line. We note that, when $\varepsilon = 0.8$, the differential privacy algorithm leads to higher availability.

In this experiment, we make $\varepsilon = 0.2$ to obtain the frequency for all nodes in the binary tree. For ease of observation, we depict four experimental results in Fig. 8. In our scheme, we regard $\varepsilon = 0.2$ as the benchmark for comparing. We note that, when $\varepsilon = 0.8$ is applied, the error is relatively small. When ε equals 0.1, big error occurs.

We can conclude that, when ε is approaching 0, the error will become large, making data unavailable. When ε is increased, the availability will be improved; however, the security and privacy protection will be weakened.

6 Conclusion

In this paper, a differential privacy protection scheme for sensitive big data in BSNs is proposed. A tree structure is constructed to reduce errors and provide long-range queries. Haar Wavelet transformation method is used to convert histogram into a complete binary tree. At last, to verify the advantages of our scheme, several experiments are conducted to show the outperformed results. Experimental results demonstrate that the tree structure greatly reduces the calculation overheads which preserving differential privacy for users.

As part of our future works, we will study how to use differential privacy scheme to protect flow data in body sensor networks.

Acknowledgments This research is sponsored in part by the National Natural Science Foundation of China (No.61402078 and 61572231). This research is also sponsored in part supported by the Fundamental Research Funds for the Central Universities (No.DUT14RC(3)090).

References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. Comput Netw 38(4):393–422
2. Alemdar H, Ersoy C (2010) Wireless sensor networks for healthcare: a survey. Comput Netw 54(15):2688–2710
3. Ali ST, Sivaraman V, Ostry D (2014) Eliminating reconciliation cost in secret key generation for body-worn health monitoring devices. IEEE Trans Mob Comput 13(12):2763–2776
4. Antonescu B, Basagni S (2013) Wireless body area networks: challenges, trends and emerging technologies. In: Proceedings of the 8th international conference on body area networks, 1-7, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)
5. Barua M, Alam MS, Liang X, Shen X (2011) Secure and quality of service assurance scheduling scheme for wban with application to ehealth. In: 2011 IEEE Wireless communications and networking conference (WCNC 2011). IEEE, pp 1102–1106
6. Bettini C, Riboni D (2015) Privacy protection in pervasive systems: State of the art and technical challenges. Pervasive Mob Comput 17:159–174
7. Canny J (2002) Collaborative filtering with privacy. In: 2002 IEEE Symposium on security and privacy. IEEE, pp 45–57
8. Casino F, Domingo-Ferrer J, Patsakis C, Puig D, Solanas A (2015) A k-anonymous approach to privacy preserving collaborative filtering. J Comput Syst Sci 81(6):1000–1011
9. Dwork C (2011) Differential privacy. In: Encyclopedia of cryptography and security. Springer, pp 338–340
10. Giannotti F, Lakshmanan LV, Monreale A, Pedreschi D, Wang H (2013) Privacy-preserving mining of association rules from outsourced transaction databases. IEEE Syst J 7(3):385–395
11. Hanson MA, Powell Jr HC, Barth AT, Ringgenberg K, Calhoun BH, Aylor JH, Lach J (2009) Body area sensor networks: challenges and opportunities. Computer 1(1):58–65
12. Hao Y, Foster R (2008) Wireless body sensor networks for health-monitoring applications. Physiol Meas 29(11):1–42
13. He D, Chen C, Chan SC, Bu J, Zhang P (2013) Secure and lightweight network admission and transmission protocol for body sensor networks. IEEE J Biomed Health Inform 17(3):664–674
14. James A (2015) Optimisation, security, privacy and trust in e-business systems. J Comput Syst Sci 81(6):941–942
15. Khan N, Javaid N, Khan ZA, Jaffar M, Rafiq U, Bibi A (2012) Ubiquitous healthcare in wireless body area networks. In: 2012 IEEE 11Th international conference on trust, security and privacy in computing and communications (trustcom 2012). IEEE, pp 1960–1967
16. Li M, Lou W, Ren K (2010) Data security and privacy in wireless body area networks. IEEE Wirel Commun 17(1):51–58
17. Li M, Yu S, Guttman JD, Lou W, Ren K (2013) Secure ad hoc trust initialization and key management in wireless body area networks. ACM Trans Sens Netw (TOSN) 9(2):1–35
18. Li N, Zhang N, Das SK, Thuraisingham B (2009) Privacy preservation in wireless sensor networks: a state-of-the-art survey. Ad Hoc Netw 7(8):1501–1514
19. Liu J, Zhang Z, Sun R, Kwak KS (2012) An efficient certificateless remote anonymous authentication scheme for wireless body area networks. In: 2012 IEEE International conference on communications (ICC 2012). IEEE, pp 3404–3408
20. Lo BP, Thiemjarus S, King R, Yang GZ (2005) Body sensor network–a wireless sensor platform for pervasive healthcare monitoring na
21. Lou H, Ma Y, Zhang F, Liu M, Shen W (2014) Data mining for privacy preserving association rules based on improved mask algorithm. In: Proceedings of the 2014 IEEE 18th international conference on computer supported cooperative work in design (CSCWD 2014). IEEE, pp 265–270
22. Lu Y, Bao SD (2014) Efficient fuzzy vault application in node recognition for securing body sensor networks. In: 2014 IEEE International conference on communications (ICC). IEEE, pp 3648–3651
23. Ma CY, Yau DK, Yip NK, Rao NS (2013) Privacy vulnerability of published anonymous mobility traces. IEEE/ACM Trans Networking 21(3):720–733
24. Mainwaring A, Culler D, Polastre J, Szewczyk R, Anderson J (2002) Wireless sensor networks for habitat monitoring. In: Proceedings of the 1st ACM international workshop on wireless sensor networks and applications. ACM, pp 88–97
25. Rushanan M, Rubin AD, Kune DF, Swanson CM (2014) Sok: security and privacy in implantable medical devices and body area networks. In: 2014 IEEE Symposium on security and privacy (SP 2014). IEEE, pp 524–539
26. Sun J, Fang Y, Zhu X (2010) Privacy and emergency response in e-healthcare leveraging wireless body sensor networks. IEEE Wirel Commun 17(1):66–73
27. Toch E, Wang Y, Cranor LF (2012) Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. User Model User-Adap Inter 22(1-2):203–220
28. Trcek D, Brodnik A (2013) Hard and soft security provisioning for computationally weak pervasive computing systems in e-health. IEEE Wirel Commun 20(4):22–29
29. Venkatasubramanian KK, Gupta SK (2010) Physiological value-based efficient usable security solutions for body sensor networks. ACM Trans Sens Netw (TOSN) 6(4):31
30. Xiao X, Wang G, Gehrke J (2011) Differential privacy via wavelet transforms. IEEE Trans Knowl Data Eng 23(8):1200–1214
31. Yan Y, Shu T (2014) Energy-efficient in-network encryption/decryption for wireless body area sensor networks. In: 2014 IEEE Global communications conference (GLOBECOM 2014). IEEE, pp 2442–2447
32. Yang GZ, Yacoub M (2006) Body sensor networks Springer

33. Yao L, Gao F, Yu G (2013) Pattern regulator for wireless body sensor networks. In: 2013 IEEE International conference on embedded and ubiquitous computing (HPCC_EUC 2013). IEEE, pp 1558–1565
34. Zhao H, Xu R, Shu M, Hu J (2015) Physiological-signal-based key negotiation protocols for body sensor networks: a survey pp 63–70. doi:10.1109/ISADS.2015.13
35. Zhou J, Cao Z, Dong X, Lin X (2015) Security and privacy in cloud-assisted wireless wearable communications: challenges, solutions, and future directions. IEEE Wirel Commun 22(2):136–144
36. Zhu T, Xiong P, Li G, Zhou W (2015) Correlated differential privacy: hiding information in non-iid data set. IEEE Trans Inf Forensics Secur 10(2):229–242