symmetry

MDPI

Article

A Practical Privacy-Preserving Publishing Mechanism Based on Personalized k-Anonymity and Temporal Differential Privacy for Wearable IoT Applications

Junqi Guo [1,2], Minghui Yang [1,*] and Boxin Wan [1]

[1] School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China; guojunqi@bnu.edu.cn (J.G.); 202021210002@mail.bnu.edu.cn (B.W.)
[2] Engineering Research Center of Intelligent Technology and Educational Application, Ministry of Education, Beijing 100875, China
* Correspondence: 201921210016@mail.bnu.edu.cn

Abstract: With the rapid development of the Internet of Things (IoT), wearable devices have become ubiquitous and interconnected in daily lives. Because wearable devices collect, transmit, and monitor humans' physiological signals, data privacy should be a concern, as well as fully protected, throughout the whole process. However, the existing privacy protection methods are insufficient. In this paper, we propose a practical privacy-preserving mechanism for physiological signals collected by intelligent wearable devices. In the data acquisition and transmission stage, we employed existing asymmetry encryption-based methods. In the data publishing stage, we proposed a new model based on the combination and optimization of k-anonymity and differential privacy. An entropy-based personalized k-anonymity algorithm is proposed to improve the performance on processing the static and long-term data. Moreover, we use the symmetry of differential privacy and propose the temporal differential privacy mechanism for real-time data to suppress the privacy leakage while updating data. It is proved theoretically that the combination of the two algorithms is reasonable. Finally, we use smart bracelets as an example to verify the performance of our mechanism. The experiment results show that personalized k-anonymity improves up to 6.25% in terms of security index compared with traditional k-anonymity, and the grouping results are more centralized. Moreover, temporal differential privacy effectively reduces the amount of information exposed, which protects the privacy of IoT-based users.

Keywords: wearable devices; IoT; k-anonymity; differential privacy; data publishing

Citation: Guo, J.; Yang, M.; Wan, B. A Practical Privacy-Preserving Publishing Mechanism Based on Personalized k-Anonymity and Temporal Differential Privacy for Wearable IoT Applications. *Symmetry* **2021**, *13*, 1043. https://doi.org/10.3390/sym13061043

Academic Editors: Weizhi Meng, Georgios Kambourakis and Jun Shao

Received: 7 May 2021
Accepted: 7 June 2021
Published: 9 June 2021

1. Introduction

Internet of things (IoT) is widely used in various fields of our daily lives, such as agriculture, industry, transportation, healthcare, education, furniture, and so on. With the IoT changing people's lifestyles, user security, in its various applications, has gradually become a problem that cannot be ignored [1,2]. In October 2016, the Mirai botnet [3] controlled a large number of IoT devices to launch DDoS attacks with network traffic up to 620 gb/s, which led to disconnections in most states in the United States. It was discovered that in 2014 there are more than 750,000 devices have been compromised and spied [4]. In 2019, a large amount of privacy problems occurred in smart-home devices, and some users' private videos were exposed on the internet by attackers. The above examples show that, due to the entity link mode of the IoT, attackers may easily obtain network data for illegal use or dissemination; more attention should be paid to this, especially in regards to the application of IoT-based health monitoring [5], since mobile health has become one of the most closely related IoT applications to consumers [6]. Data privacy of users has become a hot topic for researchers in the academic field of IoT.

An IoT infrastructure usually integrates various sensors, memories, computing units, and gateway modules to complete the tasks of data acquisition, storage, and forwarding

transmissions to a cloud platform. Compared with traditional networks, IoT device nodes are based on physical connections with smaller storage space, wider distribution, and diversified transmission protocols [7,8]. This makes common privacy protection methods difficult to be directly transplanted to IoT devices, and undoubtedly increases the risk of disclosure of privacy data, especially a large amount of physiological data, which can be used to evaluate human health status [9].

Generally, there are three basic privacy protection methods: encryption, disruption, and anonymity. Encryption is a useful method of data protection; commonly used encryption algorithms include AES [10] and RSA [11]. However, the complexity of the encryption method is also the largest of the three methods. Disruption refers to adding noise in accordance with a specific distribution of the data, which makes it difficult for attackers to accurately obtain data, such as differential privacy [12]. Anonymity is to hide some data to ensure the privacy of users, such as k-anonymity [13]. The complexities of these two methods are small, but there exist different degrees of information loss. According to the actual needs in different scenarios, the applicable methods are different.

This paper uses the intelligent wearable device (e.g., a smart bracelet) as an example to study the data privacy protection mechanism of IoT applications. Intelligent wearable devices are often attached to the skin surface to collect and analyze the physical and physiological signals of the human body. They can acquire human body status in real time as well as communicate with other IoT devices. Intelligent wearable devices are widely used in sports, healthcare, and other fields. This kind of equipment usually contains a variety of sensors, which can collect heartbeat, blood oxygen, acceleration, etc. These signals are usually transmitted to a specific gateway device (e.g., a smartphone) via the Bluetooth protocol, and then further transmitted to the server database deployed in a cloud platform. Data publishing is an important process of making data open access to provide necessary information to researchers and the public. For example, publishing infectious disease data can assist disease control centers in analyzing the diffusion trend and alert the public to take measures to prevent disease spread. Because the usability of publishing data and the low cost of data processing are basic requirements of data publishing, anonymity and disruption are common privacy protection methods in this process. Figure 1 shows the information transmission process of wearable devices, in which each stage has a certain degree of privacy leakage risk:

(1) In the signal acquisition stage, the physical structure of the equipment is at risk of being damaged.
(2) In the wireless transmission stage, the signals are faced with the risk of being intercepted by special equipment. Moreover, most IoT devices have limited computing and storage space, which makes it difficult to run complex privacy protection algorithms. For example: in a marathon race in 2014, researchers used Bluetooth sniffers to easily obtain health information from 563 different competition devices since the data collected by the devices were not protected [14].
(3) In the data publishing stage, attackers could infer the users' real information in different ways. The common methods include linkage attack [15,16] and background knowledge attack [15,17].

In order to avoid these risks, we need to develop the privacy protection technology in specific circumstances. First, it is necessary to classify the data and data types acquired by IoT devices. According to the time duration of data value maintenance, the collected data attributes can be divided into static data, long-term data, and real-time data. Considering the differences of data attributes in the application of IoT, it is essential to establish different privacy protection mechanisms for each datum attribute. Moreover, we must ensure the consistency of privacy-preserving mechanisms, which means that it cannot only be applied to a certain type of device, or weaken the effect of privacy protection when the data are updated.

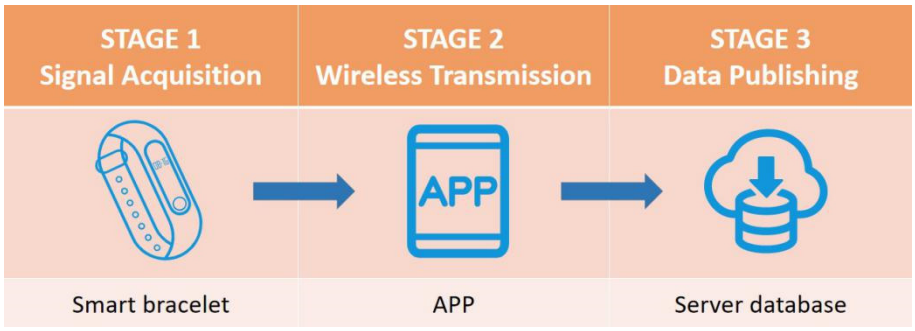


Figure 1. Information transmission process of intelligent wearable devices (e.g., a smart bracelet).

Based on the above considerations, we propose in this paper a practical privacy-preserving mechanism for wearable IoT applications. The framework is presented in Figure 2. In regards to signal acquisition during the device stage, smooth prior processing (SPA) [18] and median filtering (MF) [19] were applied to preprocess the photoplethysmography (PPG) signals collected by wearable IoT devices. Data with physical significance, such as heartbeat, blood oxygen, and acceleration were then estimated, respectively, in numerical form from the preprocessed PPG signals. In the stage involving wireless transmission from devices to the cloud, the original data were encrypted by the PRESENT algorithm [20], which is a lightweight encryption algorithm for IoT devices with limited space, and then transmitted to the server. Moreover, data were encrypted by the Paillier algorithm [21] in the cloud for homomorphic updates. In the data publishing stage, we divided the data into three parts—static data, long-term data, and real-time data, respectively. For the first and the second types of data, we designed a personalized k-anonymity algorithm to protect privacy. For the third, we proposed the temporal differential privacy mechanism to suppress the information leakage in data update. Results of the two methods will be combined and eventually released.

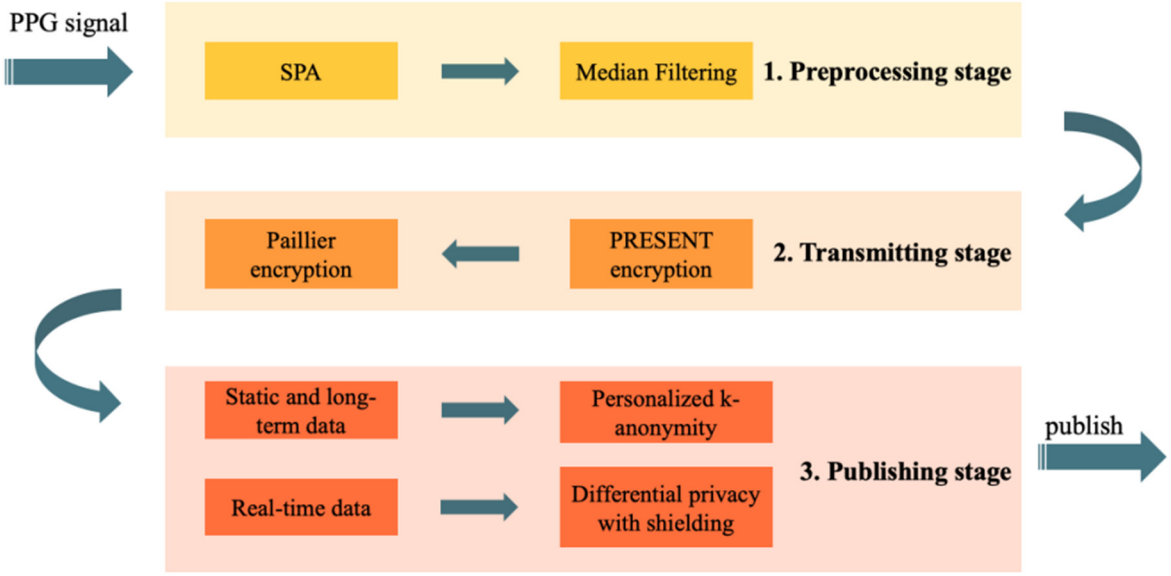


Figure 2. The framework of the proposed practical privacy-preserving mechanism in wearable IoT applications.

The contributions of this paper are mainly summarized as follows.

(1) We designed a privacy-preserving framework for IoT devices, which includes the transmitting and data publishing process.

(2) We proposed the personalized k-anonymity algorithm based on entropy of attributes to increase the usability of anonymized data, in which the category and numeric attributes are discussed as different types.
(3) We proposed the temporal differential privacy mechanism to reduce the temporal privacy disclosure, and put forward an implement algorithm in the Laplace mechanism scenarios.
(4) We proposed a practical data-publishing model for IoT devices, including the processing of static, long-term, and real-time data, and we prove that this model is of enough safety.

The remainder of our work is organized as follows. Section 2 introduces the related privacy preserving work. Section 3 presents the complete theory and method of our work, including the collection and preprocessing of data in IoT devices, the encryption methods in the transmitting process, and the publishing model containing the proposed personalized k-anonymity and temporal differential privacy. In Section 4, experiments are presented for performance evaluation and comparison. Finally, we draw conclusions in Section 5.

2. Related Work

2.1. Anonymous Methods

Definition 1. *k-anonymity [13]. The dataset **D** is grouped with at least k records and the same quasi-identifier (QI) values in each group [13].*

QI refers to the attribute or set of attributes that is different from the explicit identifier (ID, name, etc.), but can be used as the necessary evidence to fix on one specific user. In k-anonymity groups, QI remains the same for every record, and the possibility of mapping a record with a specific user can be eliminated.

Generalization [13] is one of the main methods of anonymity. This method generalizes different values to a new value at a higher level. For example, the different age values 33 and 37 in a dataset can be generalized into the new value 30–40. For more complicated problems, the generalization tree [13] is needed. In a generalization tree, every node presents the possible value at different generalizing levels. The closer a node next to the root, the higher the generalizing level it contains, and the larger the information loss in its generalizing process.

Micro-aggregation [22] is also used commonly as an anonymizing method. The records will first be divided into several groups through some heuristic algorithm, and the record values are replaced by the central record values in every group. The algorithms MD [22], MDAV [23], and V-MDAV [24] are based on this method.

Researchers have made different improvements on the k-anonymity mechanism on data level. L-diversity [15] and t-closeness [25] are proposed to prevent link attack and background knowledge attack. Razaullah Khan et al. [26] propose the θ-Sensitive k-anonymity algorithm, in which θ is set as a threshold, distinguishing different diversity level of data and adding noise. It is proved that this algorithm can prevent the sensitive variance attack and categorical similarity attack. Rohulla kosari langari et al. [27] propose a privacy preserving method KFCFA in social networks, which uses k-member fuzzy clustering for clustering and optimizes the clustering and anonymization process with the Firefly algorithm. The KFCFA method protects privacy on the data level and graph level, and effectively reduces information loss. However, the above anonymity methods lack the consideration of differences in attributes. The attributes in a dataset present heterogeneities in the data type, the value scope and the degree of dispersion, which are all important issues affecting the anonymizing results. In this paper, we propose a personalized k-anonymity algorithm to solve the problem.

2.2. Differential Privacy

Definition 2. *differential privacy [12]. If* $\boldsymbol{D}$ *and* $\boldsymbol{D}'$ *are two datasets only different in one record, they are a couple of neighbor datasets. Denote* $\boldsymbol{d}$ *the number of attributes in two datasets,* $\boldsymbol{f}$ *the function to query the dataset and returns a* $\boldsymbol{d}$*-dimension array,* $\boldsymbol{A}$ *the function to process the query result, and* ε *the setting privacy budget. Differential privacy can be described as [12]:*

$$\mathbf{Pr}[\boldsymbol{A}(\boldsymbol{D}) = \boldsymbol{O}] \leq \boldsymbol{e}^{\boldsymbol{\varepsilon}} \times \mathbf{Pr}[\boldsymbol{A}(\boldsymbol{D}') = \boldsymbol{O}] \tag{1}$$

Definition 3. *Laplace mechanism [12]. It can be proved that adding noise* $\boldsymbol{n} \sim \boldsymbol{Laplace}\left(\frac{\Delta f}{\varepsilon}\right)$ *to the query result can meet the requirement of differential privacy [12], in which* $\Delta \boldsymbol{f}$ *is the sensitivity of the query.*

In exception for a static dataset, differential privacy is also used in the publishing of temporal data. The baseline method is to add noise $\boldsymbol{n} \sim \boldsymbol{Laplace}\left(\frac{\boldsymbol{length(time\ series)}}{\varepsilon}\right)$ [28] to every time point. However, adding noise directly will cause a big information loss when the time series is long. Rastogi et al. [28] raise the $\boldsymbol{DFT_k}$ algorithm, in which the DFT coefficients of time series are perturbed by Laplace noise. Fan et al. [29] propose the Kalman algorithm, which applies Kalman filtering on the perturbed data based on time modeling. For the infinite time series, Kellaris et al. [30] raise the ω-event ϵ-differential privacy, and propose an implementation BA for the mechanism.

However, the above methods only focus on the publishing of time series of only one attribute. For practical IoT scenarios, there are data with different attributes to be published at every time. In order to preserve users' privacy, the dataset at every time point and the time series for every attribute should both be well preserved.

2.3. Privacy-Preserving in Health Data

In the aspect of health privacy protection, researchers propose various mechanisms for different scenarios.

Yinghui Zhang et al. [31] introduce a privacy-aware smart-health access control system PASH, in which attribute values of access policies are hidden in encrypted smart-health records. Hao Ren et al. [32] propose a new data aggregation scheme PMHA-DP to protect the privacy in wireless body area network, in which a privacy-enhanced average aggregation scheme (PAAS) is proposed based on differential privacy. This scheme effectively protects the user privacy in data aggregation and reduces the communication overhead. Al-Zubaidie M et al. [33] propose a new PAX authorization system to protect patient privacy in electronic health record system. The PAX system combines pseudonym, anonymity, and XACML technologies to protect privacy and reduce cost effectively.

3. Methodology

In this section, we demonstrate the details of the proposed privacy-preserving mechanism. In the application scenario of smart wearable devices, we first describe the specific ways of information collecting and processing, converting electrical signals into data. Secondly, we use two existing encryption methods to ensure the security of information in the transmitting process and background. Finally, we propose the data publishing method to reduce privacy leakage. The temporal data tables will first be split into two parts. We apply personalized k-anonymity to static and long-term data, and temporal differential privacy to real-time data. The data will be published after merging. In the end, we provide the rationality demonstration for the proposed method.

3.1. Signal Collecting and Preprocessing

In this section, we introduce the collecting and preprocessing steps of signals in smart bracelets, which are the bases of our privacy preserving mechanism. Firstly, two widely

used technologies are demonstrated for obtaining signals of heart rate and blood oxygen. Secondly, we introduce two preprocessing methods of reducing noises in signals.

3.1.1. Signal Collecting

(1) Heart rate. When the light passes through the skin tissue and then reflects to the photosensitive sensor, the absorption of light by other tissues is, basically, unchanged except for the blood, for there are blood flow changes in the artery for every beat. When the light is converted into the electrical signal in the devices, the signal can be taken as the summation of DC and AC signals, which present the unchanged signal of other tissues and the changed signal of blood flow. According to the method described in [34], we use Discrete Fourier Transform (DFT) to transform the time domain waveform of PPG signals into the frequency domain, and then extract the frequency components of human heart rate from the spectrum to obtain the heart rate data.

(2) Blood oxygen. There is a certain proportion of oxygenated hemoglobin $\boldsymbol{HbO_2}$ and hemoglobin $\boldsymbol{Hb}$ in blood. The absorption coefficient of $\boldsymbol{Hb}$ is high in the range of 600–800 nm in spectrum, while the coefficient of $\boldsymbol{HbO_2}$ is high in the range of 800–1000 nm [28]. In [28], researchers use red light (600–800 nm) and infrared ray (800–1000 nm) to detect the PPG signal of $\boldsymbol{HbO_2}$ and $\boldsymbol{Hb}$ to reflect the $\boldsymbol{SpO_2}$ value.

$$\boldsymbol{SpO_2} = \frac{\boldsymbol{C_{HbO_2}}}{\boldsymbol{C_{Hb}} + \boldsymbol{C_{HbO_2}}} \tag{2}$$

In Equation (2), $\boldsymbol{C_{HbO_2}}$ is the oxygenated hemoglobin concentration, and $\boldsymbol{C_{Hb}}$ is the reduced hemoglobin concentration.

3.1.2. Signal Preprocessing

There are other noises in the physiological signals collected by smart wearable devices, which are produced during the signal collecting process. In order to obtain the accurate physiological data of users, the signal-preprocessing module is needed in the devices.

The noises are produced because of the following two major reasons. The first is the electromyography (EMG) interference [35]. The movements of human body cause the muscle tremor, which makes the surface potential change and causes the interference of collected signals. The EMG noise is similar to the white noise, because they both present narrow in time domain and wide in frequency domain. In our work of smart bracelets, we first apply the smoothness prior approach (SPA) [18] to remove EMG noise.

The first step of SPA is dividing the signals $\boldsymbol{Z}$ into two components: the stationary part $\boldsymbol{Z_{stationary}}$ and the low-frequency aperiodic trend part $\boldsymbol{Z_{trend}}$.

$$\boldsymbol{Z} = \boldsymbol{Z_{stationary}} + \boldsymbol{Z_{trend}} \tag{3}$$

The information is contained in the stationary component, so the trend component needs to be removed. The trend component can be described as the linear model:

$$\boldsymbol{Z_{trend}} = \boldsymbol{H\theta} + v, \tag{4}$$

where $\boldsymbol{H}$ is the observation matrix, and v presents the observation error. The estimation of parameter $\widehat{\boldsymbol{\theta_\lambda}}$ is expressed as the following expression:

$$\widehat{\boldsymbol{\theta_\lambda}} = \underset{\boldsymbol{\theta}}{\mathbf{argmin}}\{\|\boldsymbol{H\theta} - \boldsymbol{Z}\|^2 + \boldsymbol{\lambda^2}\|\boldsymbol{D_d}(\boldsymbol{H\theta})\|^2\}, \tag{5}$$

where λ is the regularization parameter, and $\boldsymbol{D_d}$ is the discrete approximation to the $\boldsymbol{d}$th derivative operator. Suppose the dimension $\boldsymbol{d} = \mathbf{2}$, then the solution is:

$$\widehat{\boldsymbol{\theta_\lambda}} = \left(\boldsymbol{I} + \boldsymbol{\lambda^2}\boldsymbol{D_2}^T\boldsymbol{D_2}\right)^{-1}\boldsymbol{Z}, \tag{6}$$

and $\boldsymbol{D_2}$ is the second order difference matrix:

$$\boldsymbol{D_2} = \begin{bmatrix} \mathbf{1} & \mathbf{-2} & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{-2} & \mathbf{1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{1} & \mathbf{-2} & \mathbf{1} \end{bmatrix} \tag{7}$$

The second reason of noise interference is baseline drift, which comes from the intermittent contact problem of the devices and human body surface. In this paper, we use the median filter (MF) to suppress this kind of noise. In this filter, we set a length-fixed window and make the signals successively stream into the window. The output of the filter is the median value of all samples in the window at every time point. The points with noises that appear in isolation can be removed through MF method.

3.2. Privacy-Preserving in Data Transmitting

Encryption is the most used privacy-preserving method in the process of data transmitting, which ensures safety and causes no information loss. During the transmitting of data from devices to the cloud, we apply two different asymmetry encryption algorithms in this paper. First, in order to ensure the security of data in the process of information transmission, and the better adaption to the storage and computing space of devices, we use the lightweight PRESENT algorithm to encrypt the data inside the device. Secondly, in order to make the privacy information not exposed in the background and support the normal data update operation, we use the Paillier homomorphic encryption algorithm to encrypt the data in the cloud.

3.2.1. Encryption in Devices

In IoT devices, block cipher is widely used as a kind of encryption methods. Block cipher divides plaintext into several vectors, encrypts the vectors separately, and finally combine the ciphertexts. Block cipher decreases the size of plaintext to be encrypted at one time, which fits the limited-space IoT devices.

In the process of transmitting information from smart wearable devices to the gateway devices, we apply the PRESENT algorithm [20], which is one of the lightweight block cipher algorithms. In the PRESENT algorithm, the length of every vector is 64 bits, and the length of a secret key is 80 or 128 bits. The sizes of vectors and secret keys of PRESENT are much shorter than the traditional block cipher methods, which makes it feasible in smart wearable devices. The PRESENT algorithm is described in Figure 3.

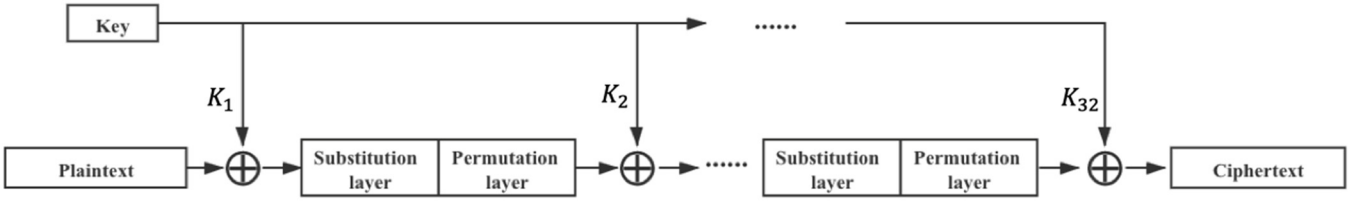


Figure 3. The encryption process of PRESENT.

In the PRESENT algorithm, each key update process round is divided into three steps [29]. Suppose that the last round key is $[k_{79}k_{78}\dots k_1k_0]$. First, the key is rotated to the left by 61 bits, in which it becomes $[k_{18}k_{17}\dots k_0k_{79}k_{78}\dots k_{20}k_{19}]$. Then, the highest four bits $k_{79}k_{78}k_{77}k_{76}$ are replaced by S-box. Finally, $k_{19}k_{18}k_{16}k_{17}k_{15}$ is exclusive-ored with the number of rounds.

The S-Box of PRESENT is shown in Table 1 [20], which is used in the substitution layer to replace every bit of text, and the function of the permutation layer [20] is calculated as

Equation (8), which means the bit i will be moved to the bit position $P(i)$. Every round of the encryption process contains the substitution layer and the permutation layer.

$$P(i) = \begin{cases} i \cdot 16 \ mod \ 63 & 0 \leq i \leq 62 \\ 63 & i = 63 \end{cases} \tag{8}$$

Table 1. The S-Box of PRESENT.

X	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
S(x)	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

3.2.2. Encryption in Cloud

When the information of users is transmitted to the cloud, in order to avoid information disclosure for background and support different computing operations, we employ the homomorphic encryption. Homomorphic encryption is a kind of encryption method, which makes the text before and after encryption homomorphic on some operations. In this paper, we use the Paillier homomorphic encryption [21] to preserve privacy in the cloud for IoT data.

The Paillier homomorphic encryption algorithm is homomorphic in both addition [21] and scalar multiplication [21]. Suppose the Paillier homomorphic encryption algorithm is f, and two plaintexts are A and B, the Paillier homomorphic expressions can be described in Equations (9) and (10).

$$f(A) + f(B) = f(A + B) \tag{9}$$

$$af(A) = f(aA). \tag{10}$$

The above two equations indicate that when data are required to be updated or computed in the cloud, we can finish the addition and scalar multiplication without decrypting, and the real data will not be exposed to untrusted platforms.

3.3. Privacy-Preserving in Data Publishing

In this part, we introduce a practical data-publishing model: for static and long-term data, the personalized k-anonymity is used, and for real-time data, the temporal differential privacy is used. We first demonstrate the data publishing in IoT in Section 3.3.1, and introduce the above two innovative algorithms in detail in Sections 3.3.2 and 3.3.3. In Section 3.3.4, we formally prove that the data-publishing model is reasonable.

3.3.1. Data Publishing of IoT

In IoT applications, analysts of relevant organizations will collect users' data for comprehensive analysis. For example, IoT developers will analyze a large amount of user data for behavior analysis and personalized services, and some useful data will be made public to support researchers' analysis in some fields. However, in the publishing process, the users' real data are published, which poses a threat to the privacy of users. If the users' private data are not properly preserved during data publishing, intentional attackers may use the information to cheat target users or sell their data, which will damage the users' interests and reduce the credibility of the IoT platforms. In addition, if the leaked information is sensitive to users, the leakage will cause psychological harm. Data that could be published in the scenarios of the smart wearable devices are listed in Table 2.

3.3.2. Personalized k-Anonymity

As we explain in Section 2, the traditional k-anonymity methods lack the consideration of attributes' differences. In this part, we introduce the personalized k-anonymity.

Table 2. Types of data collected by smart wearable devices.

Information Source	Information	Data Type	Value Duration
users	height	Numeric data	Long-term
	weight	Numeric data	Long-term
	gender	Category data	Static
	age	Numeric data	long-term
health	heart rate	Numeric data	Real-time
	blood oxygen	Numeric data	Real-time
	health level	Category data	Long-term
	diseases	Category data	Long-term
behavior	acceleration	Numeric data	Real-time
environment	city	Category data	Long-term
	GPS location	Numeric data	Real-time
	temperature	Numeric data	Real-time
	humidity	Numeric data	Real-time
	atmosphere pressure	Numeric data	Real-time

K-partition is the process of dividing an original dataset into several clusters [13], which is the most important step of k-anonymity. The common methods of k-partition are mostly based on the distance between every two records, and making the near records into the same cluster, such as MDAV [23], V-MDAV [24]. Suppose the QI attribute set is $\{q_1, q_2, \ldots, q_q\}$, the distance between two records $\{x_1, x_2, \ldots, x_q\}$ and $\{y_1, y_2, \ldots, y_q\}$ can be calculated as:

$$distance = \sqrt{\omega_1(x_1 - y_1)^2 + \omega_2(x_2 - y_2)^2 + \ldots + \omega_q(x_q - y_q)^2} \tag{11}$$

where $\{\omega_i\}$ are weights of every attribute. In k-partition, we notice that there are two important points: one is the distance of values for one attribute $|x_i - y_i|$, the other is the assigned weight ω_i of every attribute.

Firstly, we define the distance of two values of an attribute. As for the digital attributes, we use the absolute value of subtraction. As for the category attributes, we use the step to the same parent node in the generalization tree to measure the distance between the two child nodes.

To assign the weights, the entropy weight method [36] is used in this paper. We describe the reason of using this method in a simple way: the attributes of high entropy are always complicated in the occurrence of value, which should play more important roles in dividing data into clusters. As what have discussed in the previous sections, the discussion of entropy weight method also bases on category and numeric attributes. As for the category attributes, the method is to find frequency f_{ij} for each possible value i, usually f_{ij} is the ratio of the number of occurrences of the value i and the number of records. The normalized entropy of attribute j is:

$$En_j = -\frac{1}{\ln p}\sum_{i=1}^{p} f_{ij} \ln f_{ij} \tag{12}$$

As for the numeric data, the scope of each attribute value is different, which will affect the extent of dispersion of attribute. Data should be normalized before assigning weights to attributes, so that we can judge the dispersion degree of each attribute from a unified perspective. In our study, we use the membership degree [37] method to normalize the values.

The data collected by devices in one time can be described as $D_{p\times q}$, where p presents the numbers of users, and q presents the numbers of QIs. The membership function can be expressed as:

$$\mu_{ij} = \begin{cases} 1, D_{i,j} \leq D_{j1} \\ \frac{D_{j2}-D_{i,j}}{D_{j2}-D_{j1}}, D_{j1} < D_{i,j} < D_{j2} \\ 0, \ D_{i,j} \geq D_{j2} \end{cases} \tag{13}$$

D_{j1} and D_{j2} are the upper limits of satisfying value and permissible value of attribute q_j. The normalized matrix can be expressed as $\mu_{p\times q}$. Each element in the matrix ranges from 0 to 1. According to the idea we put forward, the greater the dispersion of data, the bigger the entropy of the data, and the less security the data possess. Entropy of an attribute q_j is calculated as in Equation (12), in which

$$f_{ij} = \frac{\mu_{ij}}{\sum_{i=1}^{p} \mu_{ij}}. \tag{14}$$

The weight assigned to each attribute is:

$$\omega_j = \frac{1 - En_j}{q - \sum_{j=1}^{q} En_j} \tag{15}$$

However, the cost time for computing the weight of every attribute is almost equal to the cost time of k-partition. For the occasions where there are a large amount of data, the process to assign weights will cost much time. When we design the personalized k-anonymity algorithm, this procedure is based on a small amount of data sampled randomly to reduce the time cost. It can be proved that the result of weight is the unbiased estimation of the complete data result, for the calculating of entropy is based on frequency.

We use the V-MDAV algorithm [24] for k-anonymity grouping, which is presented in Algorithm 1. The personalized k-anonymity algorithm is described in Algorithm 2.

Algorithm 1. V-MDAV

Input: distance matrix ***DM***, Parameter k
Output: micro-aggregated set ***M***
1: $c = compute\ centroid\ record(\boldsymbol{DM})$
2: while (more than $k-1$ records wait to be assigned) do
3: $\quad e = the\ most\ distant\ record\ to\ c$
4: $\quad g_i = build\ group\ from\ record(e, \boldsymbol{DM}, k)$
5: $\quad g_i = extend\ the\ group(g_i, \boldsymbol{DM}, k)$
6: end while
7: $g_1, g_2, \ldots, g_s = assign\ remaining\ records(\boldsymbol{DM}, g_1, g_2, \ldots, g_s)$
8: $\boldsymbol{M} = build\ microaggregated\ set(g_1, g_2, \ldots, g_s)$
9: return ***M***
10: end function

3.3.3. Temporal Differential Privacy

Apart from the common insecurity that a statistic dataset will bring, dangers of data leaking in a temporal process still exist. For example, Table 3 shows the information that a device collects at different moments from one user. This user's heart rate increases suddenly, while the other two attributes do not change a lot. Moreover, the background attackers can draw a conclusion that it is of a low probability that the sudden increase in the heart rate is related to sports, but probably results from an illness, such as the sudden palpitation. If this user really has the illness, the health information is exposed to the attackers.

Algorithm 2. Personalized k-anonymity

Input: Original datasets $\boldsymbol{D}$, Parameter k, sampling ratio s
Output: published datasets $\boldsymbol{PD}$
$\boldsymbol{SD} = random\ sampling(\boldsymbol{D},\ s)$
while $j \leq attribute\ number(\boldsymbol{SD})$
    if $attribute(j)$ *is numeric*
        $\boldsymbol{\mu} = membership(\boldsymbol{SD}(j))$
        $\boldsymbol{f}_{ij} = \frac{\boldsymbol{\mu}_{ij}}{\sum_{i=1}^{p} \boldsymbol{\mu}_{ij}}$
    else if $attribute(j)$ *is category*
        $\boldsymbol{f}_{ij} = \frac{occurrences(i)}{record\ number(\boldsymbol{SD})}$
    end if
    $\boldsymbol{En}_j = -\frac{1}{\ln p} \sum_{i=1}^{p} \boldsymbol{f}_{ij} \ln \boldsymbol{f}_{ij}$
end while
$\boldsymbol{\omega} = weight\ assigning(\boldsymbol{En})$
$\boldsymbol{DM} = compute\ distance\ matrix(\boldsymbol{\omega},\ \boldsymbol{D})$
$\boldsymbol{M} = V - MDAV(\boldsymbol{DM},\ k)$
$\boldsymbol{PD} = anonymity(\boldsymbol{M})$

Table 3. Information that device collects at two different moments from one user.

Time	SpO_2 (%)	Heartbeat (bpm)	Steps Increase
10:20:43	99	80	0
10:22:18	99	102	2

In a dataset, some health attributes may have common correlations, which can be positive, negative, or more complex. Therefore, unbalanced changes of attributes indicate the occurrence of abnormal conditions and can lead to information leakage.

In this paper, we propose the temporal differential privacy mechanism to solve the above problem. We firstly define an important variable δ to indicate the sensitivity ratio between two time points. Suppose the sensitivities at time 0 and time t are Δf and Δf_t, then the sensitivity ratio δ is $\frac{\Delta f_t}{\Delta f}$.

Definition 4. *Temporal differential privacy. Suppose the range of* $\frac{Pr[A(\boldsymbol{D})=\boldsymbol{O}]}{Pr[A(\boldsymbol{D_t})=\delta\boldsymbol{O}]}$ *is* $[r_1,\ r_2]$*, the published results are satisfied with temporal differential privacy when* $r_1 \leq 1 \leq r_2$*.*

In this paper, we put forward an implementation method of temporal differential privacy in the Laplace mechanism scenarios.

Suppose the original query result is $f(\boldsymbol{D}) = (x_1,\ x_2,\ \ldots, x_d)^T$, and the result at time t is $f(\boldsymbol{D_t}) = (x_{t1},\ x_{t2},\ \ldots, x_{td})^T = (x_1 + \Delta x_{t1},\ x_2 + \Delta x_{t2},\ \ldots, x_d + \Delta x_{td})^T$. According to the Laplace mechanism [11], the result perturbed by Laplace noise $\boldsymbol{n} \sim \boldsymbol{Laplace}\left(\frac{\Delta f}{\varepsilon}\right)$ is satisfied with differential privacy. From the definition of Laplace distribution, the distribution of published result at time 0 of differential privacy is

$$\Pr[A(\boldsymbol{D}) = \boldsymbol{O}] = \prod_{i=1}^{d} \frac{\varepsilon}{2\Delta f} e^{-\frac{\varepsilon}{\Delta f}|y_i|}. \tag{16}$$

At time t, the published result is also perturbed by Laplace noise $\boldsymbol{n} \sim \boldsymbol{Laplace}\left(\frac{\Delta f_t}{\varepsilon}\right)$, and a distribution also exists. We study the possible result $\boldsymbol{O_t} = \delta\boldsymbol{O} = (\delta y_1,\ \delta y_2,\ \ldots, \delta y_d)^T$. Through the same way as above, the probability is

$$\Pr[A(\boldsymbol{D_t}) = \boldsymbol{O_t}] = \prod_{i=1}^{d} \frac{\varepsilon}{2\Delta f_t} e^{-\frac{\varepsilon}{\Delta f_t}|\Delta x_{ti} - \delta y_i|}. \tag{17}$$

In order to meet the requirement of temporal differential privacy, we first compute the result of $\frac{\Pr[A(\boldsymbol{D})=\boldsymbol{O}]}{\Pr[A(\boldsymbol{D_t})=\delta\boldsymbol{O}]}$ in Inequality 18. We use the absolute value inequality at the position of less-than-equal sign.

$$\begin{aligned}\frac{\Pr[A(\boldsymbol{D})=\boldsymbol{O}]}{\Pr[A(\boldsymbol{D_t})=\boldsymbol{O_t}]} &= \delta^d \exp\left\{\varepsilon \sum_{i=1}^{d} \frac{|\Delta x_{ti}-\delta y_i|}{\Delta f_t} - \frac{|y_i|}{\Delta f}\right\} \\ = \delta^d \exp\left\{\varepsilon \sum_{i=1}^{d} \frac{|\Delta x_{ti}-\delta y_i|-\delta|y_i|}{\delta\Delta f}\right\} &\leq \delta^d \exp\left\{\frac{\varepsilon}{\delta} \sum_{i=1}^{d} \frac{|\Delta x_{ti}|}{\Delta f}\right\} = \delta^d \exp\left\{\frac{\varepsilon}{\Delta f_t} \sum_{i=1}^{d} |\Delta x_{ti}|\right\}\end{aligned} \tag{18}$$

The computing result shows that the upper bound is not of correlations with the possible result $\boldsymbol{O}$, which means this inequality is true for any value of $\boldsymbol{O}$. Moreover, from the symmetry of $\Pr[A(\boldsymbol{D}) = \boldsymbol{O}]$ and $\Pr[A(\boldsymbol{D_t}) = \boldsymbol{O_t}]$, the two bounds of $\frac{\Pr[A(\boldsymbol{D})=\boldsymbol{O}]}{\Pr[A(\boldsymbol{D_t})=\boldsymbol{O_t}]}$ can also be computed:

$$\delta^d e^{-\frac{\varepsilon}{\Delta f_t}\sum_{i=1}^{d}|\Delta x_{ti}|} \leq \frac{\Pr[A(\boldsymbol{D}) = \boldsymbol{O}]}{\Pr[A(\boldsymbol{D_t}) = \boldsymbol{O_t}]} \leq \delta^d e^{\frac{\varepsilon}{\Delta f_t}\sum_{i=1}^{d}|\Delta x_{ti}|} \tag{19}$$

According to definition 4, temporal differential privacy is satisfied if

$$\begin{cases} \delta^d e^{\frac{\varepsilon}{\Delta f_t}\sum_{i=1}^{d}|\Delta x_{ti}|} > 1 \\ \delta^d e^{-\frac{\varepsilon}{\Delta f_t}\sum_{i=1}^{d}|\Delta x_{ti}|} < 1 \end{cases}. \tag{20}$$

We take the logarithm of both sides, and then merge the two intermediate results together to compute the final result:

$$\frac{\sum_{i=1}^{d}|\Delta x_{ti}|}{\Delta f_t|\ln\delta|} > \frac{d}{\varepsilon}. \tag{21}$$

We take Inequality 21 as the condition of our algorithm, which represents the bound of whether the data can be published, ensuring the data will not leak important information in the variations. The above algorithm is described in Algorithm 3. It is obvious that this algorithm is satisfied with temporal differential privacy.

Algorithm 3. Temporal differential privacy

Input: Original datasets $\{\boldsymbol{T}_0, \boldsymbol{T}_1, \ldots, \boldsymbol{T}_m\}$, ε
Output: Processed results $\{\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_n\}$, $n \leq m$
1: $\Delta f = sensitivity(\boldsymbol{T}_0)$, $k = 0$, $j = 0$
2: while $k \leq m$ do
3: $\quad\Delta f_k = sensitivity(\boldsymbol{T}_k)$
4: $\quad\delta = \frac{\Delta f_k}{\Delta f}$
5: $\quad$if $\frac{\sum_{i=1}^{d}|\Delta x_{ki}|}{\Delta f_k|\ln\delta|} > \frac{d}{\varepsilon}$
6: $\quad\quad$return $\boldsymbol{R}_j = differential\ privacy(\varepsilon, \boldsymbol{T}_k)$
7: $\quad\quad\Delta f = \Delta f_k$, $k \leftarrow k+1$, $j \leftarrow j+1$
8: $\quad$end if
9: end while

3.3.4. Rationality Demonstration

We proposed a method that combines the promotions of the two traditional algorithms: k-anonymity and differential privacy in data publishing for different kinds of attributes in Sections 3.3.2 and 3.3.3. In this section, we prove that this combination is rational and the final result satisfies with both k-anonymity and the differential privacy mechanism.

The real-time attributes can be seen as the sensitive attributes (SA), while the static and long-term attributes set includes QIs and SAs. Suppose the static and long-term attributes set is $(QI,\ SA_1)$, and the real-time attributes set is SA_2.

(1) K-anonymity. The values of QI are the same in an equivalence class as the attributes $(QI,\ SA_1)$ according to the k-anonymity mechanism. It is obvious that, as for the combined attributes set, $(QI,\ SA_1,\ SA_2)$, the QI values remain the same in the equivalence class as the record number at least k. Therefore, our combined mechanism is satisfied with k-anonymity.

(2) Differential privacy. According to the differential privacy mechanism, the relationship between neighbor datasets $\boldsymbol{SA_2}$ and $\boldsymbol{SA_2}'$ satisfies with the expression:

$$\Pr[A(\boldsymbol{SA_2}) = \boldsymbol{O}_1] \leq e^{\varepsilon} \times \Pr\left[A\left(\boldsymbol{SA_2}'\right) = \boldsymbol{O}_1\right], \tag{22}$$

Other parameters are the same as Section 2.

$\boldsymbol{SA_2}$ and $\boldsymbol{SA_2}'$ are a couple of neighbor datasets, which means that they are different in only one record. We add the same data $(\boldsymbol{QI},\ \boldsymbol{SA_1})$ that have been anonymized to the two datasets. It is obvious that the new datasets $\boldsymbol{D}(\boldsymbol{QI},\ \boldsymbol{SA_1},\ \boldsymbol{SA_2})$ and $\boldsymbol{D}'(\boldsymbol{QI},\ \boldsymbol{SA_1},\ \boldsymbol{SA_2}')$ are neighbor datasets to each other.

Denote PM the publishing mechanism in this paper and $\boldsymbol{O} = (\boldsymbol{O_1},\ \boldsymbol{O_2})$, and we can conclude from the above that

$$\Pr[PM(\boldsymbol{D}) = \boldsymbol{O}] \leq e^{\varepsilon} \times \Pr\left[PM\left(\boldsymbol{D}'\right) = \boldsymbol{O}\right]. \tag{23}$$

It is proved that the proposed data publishing mechanism is satisfied with the differential privacy.

4. Results and Discussion

In this section, we evaluate the effectiveness of the data publishing algorithms proposed in Section 3, including personalized k-anonymity and temporal differential privacy. We compare the performances of the proposed algorithm with the existing algorithms by information loss, distance linked closure risk, and other quantitative results.

4.1. Dataset

The data in the following experiments were collected from 513 teenagers in a middle school using our smart bracelets. The dataset contains the heart rate (HR), blood oxygen (SpO_2), and other physiological data before and after a long run.

The attributes can be divided into three kinds: removed data, static and long-term data, and real-time data, which are listed in detail as follows:

(1) Removed data. The removed attribute set contains names and IDs of devices, which will be removed before applying our privacy preserving method.
(2) Static and long-term data. The static and long-term attribute set contains gender, age, height, and the health level of students, in which QI set consists of {gender, age, height}, and SA set consists of {health level}.
(3) Real-time data. The real-time attribute set contains the resting HR, descent rate of HR, increase rate of HR, HR reserve, SpO_2 saturation mean, SpO_2 saturation standard deviation, HR after exercise and exercise time duration.

Note that the following experiments are implemented in the Python 3.7.4 development environment. At the beginning of our experiment, we removed the names and IDs in the dataset first. Then, we apply personalized k-anonymity mentioned in Section 3.3.2 to the static and long-term data, and temporal differential privacy mentioned in Section 3.3.3 to the real-time data.

4.2. Privacy Preserving on Static Data

We process the static and long-term data with a personalized k-anonymity publishing scheme propose in Section 3.3.2, which is compared with the conventional V-MDAV. In

this part, we use two indexes to evaluate the usability and safety of the two algorithms.

We use information loss (IL) [24] to measure the usability of data. Suppose $\boldsymbol{D}$ is a dataset, and $\boldsymbol{D}'(g_1, g_2, \ldots, g_m)$ is the dataset to be published, $\{g_1, g_2, \ldots, g_m\}$ are groups that are divided by k-anonymity. The group square errors (GSE) of g_i can be calculated as:

$$GSE(g_i) = \sum_{i=1}^{n_i} distance\left(g_{ij}, \overline{g_i}\right), \tag{24}$$

where $\overline{D_i}$ is the centroid of g_i, and n_i is the number of records in group g_i. Moreover, the sum of square errors (SSE) of $\boldsymbol{D}'$ is defined as:

$$SSE = \sum_{i=1}^{m} GSE \tag{25}$$

The total sum of squares (SST) is defined as:

$$SST = \sum_{i=1}^{p} distance\left(\boldsymbol{D}_i, \overline{\boldsymbol{D}}\right), \tag{26}$$

where $\overline{\boldsymbol{D}}$ is the average of total data in $\boldsymbol{D}$.

IL is the ratio of SSE and SST, which is expressed in Equation (27).

$$IL = \frac{SSE}{SST} \tag{27}$$

We use the distance linked disclosure risk (DLD) [38] of the anonymized dataset to measure the disclosure risk of k-anonymity. DLD is calculated as:

$$DLD = \frac{linked - record - num}{total - record - num}, \tag{28}$$

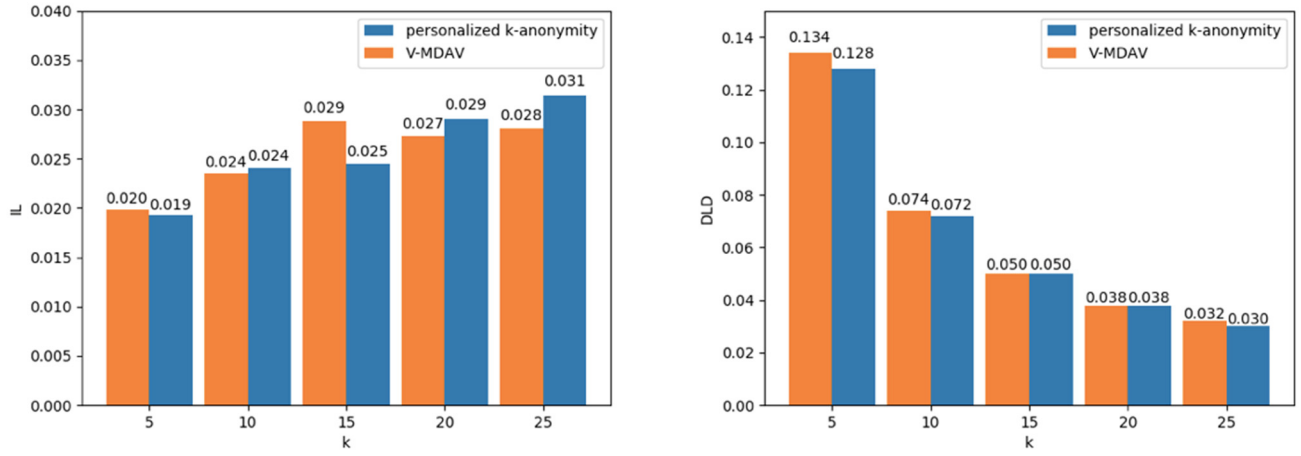
where $linked - record - num$ is the number of linked records, and $total - record - num$ is the number of total records of the anonymized data. A smaller DLD means the anonymized table takes less risk of information leakage.

Table 4 presents the results of calculated weights of QIs based on the personalized k-anonymity. For different values of k, assessments of V-MDAV and personalized k-anonymity are shown in Figure 4, and Table 5 is the group information of the two algorithms.

Table 4. Weight of QIs.

Attribute	Weight
gender	0.34497
age	0.02572
height	0.62931

From Figure 4, we can conclude that when the value of k is small, the information loss of personalized k-anonymity is smaller. When k value is large, the information loss of V-MDAV is smaller. Therefore, personalized k-anonymity has better usability in the case of less k value and more groups. Overall, the IL values of the two algorithms are almost equal for different values of k. However, our algorithm improves up to 6.25% in terms of security index compared with traditional k-anonymity. Moreover, Table 5 shows that the group size variance of personalized k-anonymity is smaller than V-MDAV, and there are more anonymous groups in the results of our algorithm, which means the personalized k-anonymity divides groups more evenly, preventing too large or too small groups, which can improve the usability for published data.



(**a**) (**b**)

Figure 4. IL (**a**) and DLD (**b**) of the results of V-MDAV and personalized k-anonymity.

Table 5. Group information of V-MDAV and personalized k-anonymity.

k	Algorithm	Group Number	Maximum Group Size	Group Size Variance
$k = 10$	V-MDAV	36	22	9.099
	personalized k-anonymity	37	18	4.938
$k = 20$	V-MDAV	19	32	10.825
	personalized k-anonymity	19	31	9.407

4.3. Privacy Preserving on Real-Time Data

In the experiment for privacy preserving on real-time data, we implement the temporary differential privacy and the traditional differential privacy mechanisms, and compare their performances through the complexity of time series.

We suppose the query f is the function to calculate the maximum value of every attribute in the dataset, and returns the data in the form of $\{x_1, x_2, \ldots, x_d\}$. In the following experiments, we set $\varepsilon = 0.5$ and test 100 time points.

The approximate entropy ($ApEn$) [39] indicates the complexity of a time series. A bigger $ApEn$ means that the time series contains more information. $ApEn$ of a time series $\boldsymbol{u} = [u(1), u(2), \ldots, u(N)]$ is computed in the light of the following steps [39]:

Step 1. Decide the parameters m and r. m is an integer of the length of array in **Step 2**, which is at least 2. r is a real number representing the measure of similarity of time series. In common cases, we set

$$r = 0.2 * std, \tag{29}$$

where std is the standard deviation of $\boldsymbol{u}$.

Step 2. Reconstruct m-dimension vectors of continuous time: $\boldsymbol{X}(1), \boldsymbol{X}(2), \ldots, \boldsymbol{X}(N-m+1)$, where

$$\boldsymbol{X}(i) = [u(i), u(i+1), \ldots, u(i+m-1)]. \tag{30}$$

Step 3. For $1 \leq i \leq N-m+1$, compute the proportion $C_i^m(r)$ of similar vectors to $\boldsymbol{X}(i)$.

$$C_i^m(r) = \frac{numbers\ of\ \boldsymbol{X}(j)\ that\ d[\boldsymbol{X}(i), \boldsymbol{X}(j)] \leq r}{N-m+1}, \tag{31}$$

where $d[\boldsymbol{X}, \boldsymbol{X}^*] = \max_a |u(a) - u^*(a)|$, and $1 \leq j \leq N-m+1$.

Step 4. Compute the entropy $\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\log(C_i^m(\mathrm{r}))$. Repeat the above steps to compute $\Phi^{m+1}(r)$, and the $ApEn$ of time series $\boldsymbol{u}$ is

$$ApEn = \Phi^m(r) - \Phi^{m+1}(r) \tag{32}$$

Figure 5 shows the $ApEn$s of the publishing results of differential privacy and temporal differential privacy over time. In Figure 5, we find that the temporal differential privacy reduces about 12% amount of information based on the traditional differential privacy. Moreover, the complexity of the temporal differential privacy result is both lower than the original data and the differential privacy result at some time, which means the reduced information must contain the information from the original data. The detailed results of temporal differential privacy for different ε and queries are presented in Table 6.

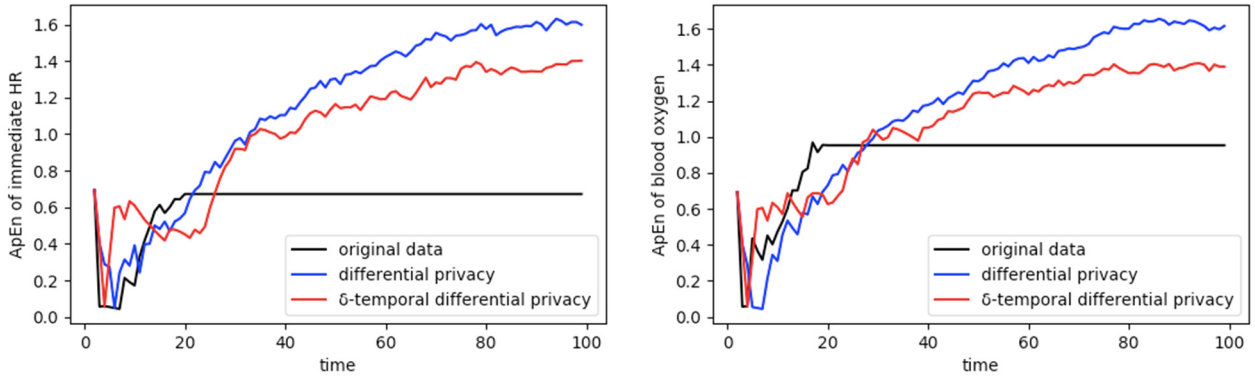


Figure 5. $ApEn$s of attributes in differential privacy and temporal differential privacy.

Table 6. Results of temporal differential privacy for different ε and queries.

ε	Query	Number of Published Records	Average *ApEn* of Temporal Differential Privacy	Average *ApEn* of Differential Privacy
$\varepsilon = 0.2$	maximum	13	0.3514	1.6063
	average	22	0.5035	1.6980
$\varepsilon = 0.5$	maximum	18	0.4893	1.5504
	average	52	1.1296	1.6780

5. Conclusions and Future Work

In this paper, we propose a practical privacy-preserving mechanism to ensure data security in different stages of a wearable IoT framework, in which the application of smart wearable devices is taken as an example. First, we employ the light-weighted PRESENT algorithm to encrypt information in IoT devices, and utilize Paillier homomorphic encryption to manage data on the cloud platform. In the publishing data stage, we optimize the traditional *k*-anonymity algorithm for static data and the differential privacy algorithm for the real-time data, and then make the rational demonstration for the combination of the two optimized algorithms. Specifically, we propose the personalized *k*-anonymity algorithm, in which we assign weights to different attributes, based on the entropy, and discuss the different occasions for the numeric data and category data. The experiment results show that personalized k-anonymity is equivalent to traditional k-anonymity in usability, but its safety index is about 0–6.25% higher than traditional algorithms. Moreover, its grouping results are more concentrated. Furthermore, we propose the temporal differential privacy mechanism to ensure the privacy security in temporal dataset, and put forward an implementing method based on the Laplace mechanism. The experiment results show that the temporal differential privacy decreases the disclosure in time variation duration.

Taken together, our results provide evidences toward the feasibility and effectiveness of our mechanism in protecting privacy for IoT-based users.

In the future work, we will improve the proposed mechanism from the following aspects:

(1) There are some researches of attacks on the PRESENT algorithm, such as [40]. We will improve the algorithm in future work to enhance security.
(2) Some existing smart bracelet systems have used learning algorithms for classifying and predicting tasks, for example, the health status of users could be evaluated according to the data collected by the smart bracelets. In the training process, users' privacy will also be exposed. We intend to adopt the federal learning method in the future work.
(3) We will improve our mechanism to adapt to other kinds of IoT devices, and evaluate its effectiveness in the current network and device environment.

Author Contributions: Conceptualization, J.G.; methodology, J.G. and M.Y.; software, M.Y.; validation, J.G., M.Y., and B.W.; formal analysis, M.Y. and B.W.; investigation, J.G. and M.Y.; resources, J.G. and M.Y.; data curation, M.Y. and B.W.; writing—original draft preparation, J.G. and M.Y.; writing—review and editing, J.G. and B.W.; visualization, M.Y. and B.W.; supervision, J.G.; project administration, J.G.; funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research is sponsored by National Natural Science Foundation of China (No.61977006), "Educational Big Data R&D and its Application"—Major Big Data Engineering Project of National Development and Reform Commission 2017, and Beijing Advanced Innovation Center for Future Education (BJAICFE2016IR-004).

Data Availability Statement: The data used to support the findings of this study have not been made available due to the private information of teenagers.

Conflicts of Interest: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work.

References

1. Nour, B.; Sharif, K.; Li, F.; Biswas, S.; Moungla, H.; Guizani, M.; Wang, Y. A survey of Internet of Things communication using ICN: A use case perspective. *Comput. Commun.* **2019**, *142*, 95–123. [CrossRef]
2. Ala, A.F.; Mohsen, G.; Mehdi, M.; Mohammed, A.; Moussa, A. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2347–2376. [CrossRef]
3. Zhang, X.L.; Upton, O.; Beebe, N.L.; Choo, R.K.K. IoT Botnet Forensics: A Comprehensive Digital Forensic Case Study on Mirai Botnet Servers. *Forensic Sci. Int. Digit. Investig.* **2020**, *32*, 300926. [CrossRef]
4. Kumar, P.; Braeken, A.; Gurtov, A.; Iinatti, J.; Ha, P.H. Anonymous Secure Framework in Connected Smart Home Environments. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 968–979. [CrossRef]
5. Ghosh, A.; Raha, A.; Mukherjee, A. Energy-Efficient IoT-Health Monitoring System using Approximate Computing. *Internet Things* **2020**, *9*, 100166. [CrossRef]
6. Marcus, A.G.; Santos, R.M.; Rodrigo, O.; Petro, P.R.F.; Javier, D.S.; Victor, H.C.d.A. Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. *Inf. Fusion* **2020**, *53*, 222–239. [CrossRef]
7. Feroz Khan, A.B.; Anandharaj, G. A cognitive key management technique for energy efficiency and scalability in securing the sensor nodes in the IoT environment. CKMT. *SN Appl. Sci.* **2019**, *1*, 1–7. [CrossRef]
8. Rafik, H.; Zheng, Y.; Khan, M. A privacy-preserving cryptosystem for IoT E-healthcare. *Inf. Sci.* **2019**, *527*, 493–510. [CrossRef]
9. Ojetunde, B.; Shibata, N.; Gao, J.T. Monitoring-Based Method for Securing Link State Routing against Byzantine Attacks in Wireless Networks. *J. Inf. Process.* **2018**, *26*, 98–110. [CrossRef]
10. National Institute of Standards and Technology (NIST). Advanced Encryption Standard. NIST, FIPS PUB 197, US Department of Commerce. 2001. Available online: http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf (accessed on 6 May 2021).
11. Rivest, R.L.; Shamir, A.; Adleman, L. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM* **1978**, *21*, 120–126. [CrossRef]
12. Dwork, C. Differential Privacy. In Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, Venice, Italy, 10–14 July 2006.
13. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **2002**, *10*, 557–570. [CrossRef]
14. He, D.; Chan, S.; Guizani, M. User privacy and data trustworthiness in mobile crowd sensing. *IEEE Wirel. Commun.* **2015**, *22*, 28–34. [CrossRef]

15. Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data (TKDD)* **2007**, *1*, 3. [CrossRef]
16. Chen, R.; Fung, B.C.M.; Mohammed, N.; Desai, B.C.; Wang, K. Privacy-preserving trajectory data publishing by local suppression. *Inf. Sci.* **2013**, *231*, 83–97. [CrossRef]
17. Pan, X.; Chen, W.Z.; Wu, L. Mobile User Location Inference Attacks Fusing with Multiple Background Knowledge in Location-Based Social Networks. *Mathematics* **2020**, *8*, 262. [CrossRef]
18. Zhang, F.; Chen, S.; Zhang, H.; Zhang, X.; Li, G. Bioelectric signal detrending using smoothness prior approach. *Med. Eng. Phys.* **2014**, *36*, 1007–1013. [CrossRef]
19. Chen, T.; Wu, H.R. Adaptive impulse detection using center-weighted median filters. *IEEE Signal Process. Lett.* **2001**, *8*, 1–3. [CrossRef]
20. Leander, G.; Paar, C.; Poschmann, A.; Robshaw, M.J.B.; Seurin, Y.; Bogdanov, A.; Knudsen, L.R.; Vikkelsoe, C. PRESENT: An Ultra-Lightweight Block Cipher. In Proceedings of the 9th International Workshop CHES 2007, Vienna, Austria, 10–13 September 2007; Springer: Berlin, Germany, 2007.
21. Tsoutsos, N.G.; Maniatakos, M. The HEROIC framework: Encrypted computation without shared keys. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **2015**, *34*, 875–888. [CrossRef]
22. Domingo-Ferrer, J.; Mateo-Sanz, J.M. Practical dataoriented microaggregation for statistical disclosure control. *IEEE Trans. Knowl. Data Eng.* **2002**, *14*, 189–201. [CrossRef]
23. Domingo-Ferrer, J.; Torra, V. Ordinal, continuous and heterogeneous k-anonymity through micro-aggregation. *J. Data Min. Knowl. Discov. Sep.* **2005**, *11*, 195–202. [CrossRef]
24. Solanas, A.; Martinez-Ballesté, A. V-MDAV: A multivariate microaggregation with variable group size. In Proceedings of the Seventh Compstat Symposium of the Lasc, Rome, Italy, 28 August–1 September 2006.
25. Li, N.; Li, T.; Venkatasubramanian, S. T-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering, Istanbul, Turkey, 15–20 April 2007.
26. Khan, R.; Tao, X.; Anjum, A.; Kanwal, T.; Malik, S.u.R.; Khan, A.; Rehman, W.u.; Maple, C. θ-Sensitive k-Anonymity: An Anonymization Model for IoT based Electronic Health Records. *Electronics* **2020**, *9*, 716. [CrossRef]
27. Langari, R.K.; Sardar, S.; Mousavi, S.A.A.; Radfar, R. Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks. *Expert Syst. Appl.* **2020**, *141*, 112968. [CrossRef]
28. Rastogi, V.; Nath, S. Differentially private aggregation of distributed time-series with transformation and encryption. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Indianapolis, IN, USA, 6–10 June 2010; pp. 735–746.
29. Fan, L.Y.; Xiong, L.; Sunderam, V. Differentially private multi-dimensional time series release for traffic monitoring. In Proceedings of the 27th international conference on Data and Applications Security and Privacy XXVII, Newark, NJ, USA, 15–17 July 2013; pp. 33–48.
30. Kellaris, G.; Papadopoulos, S.; Xiao, X.K.; Papadias, D. Differentially private event sequences over infinite streams. *Proc. Vldb Endow.* **2014**, *7*, 1155–1166. [CrossRef]
31. Zhang, Y.; Zheng, D.; Deng, R.H. Security and Privacy in Smart Health: Efficient Policy-Hiding Attribute-Based Access Control. *IEEE Internet Things J.* **2018**, *5*, 2130–2145. [CrossRef]
32. Ren, H.; Li, H.; Liang, X.; He, S.; Dai, Y.; Zhao, L. Privacy-Enhanced and Multifunctional Health Data Aggregation under Differential Privacy Guarantees. *Sensors* **2016**, *16*, 1463. [CrossRef]
33. Al-Zubaidie, M.; Zhang, Z.; Zhang, J. PAX: Using Pseudonymization and Anonymization to Protect Patients' Identities and Data in the Healthcare System. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1490. [CrossRef]
34. Saquib, N.; Papon, M.T.I.; Ahmad, I.; Rahman, A. Measurement of heart rate using photoplethysmography. In Proceedings of the 2015 International Conference on Networking Systems and Security, Dhaka, Bangladesh, 5–7 January 2015; pp. 1–6.
35. Zhang, Z. Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1902–1910. [CrossRef]
36. Xie, H.; Cheng, H.Z.; Niu, D.X. Discretization Algorithm for Continuous Attributes of Rough Sets Based on Information Entropy. *Chin. J. Comput.* **2005**, *28*, 1570–1574.
37. Sancho-Royo, A.; Verdegay, J.L. Methods for the construction of membership functions. *Int. J. Intell. Syst.* **1999**, *14*, 1213–1230. [CrossRef]
38. Domingo, F.J. Microaggregation for Database and Location Privacy. In Proceedings of the 6th International Conference, NGITS 2006, Kibbutz Shefayim, Israel, 4–6 July 2006.
39. Kanungo, T.; Mount, D.M.; Netanyahu, N.S.; Piatko, C.; Silverman, R.; Wu, A.Y. Computing Nearest Neighbors for Moving Points and Applications to Clustering. In Proceedings of the 10th Annual ACM-SIAM Symp. Discrete Algorithms. Omni Inner Harbor Hotel, Baltimore, MD, USA, 17–19 January 1999; pp. 931–932.
40. Duan, X.; Cui, Q.; Wang, S.; Fang, H.; She, G. Differential power analysis attack and efficient countermeasures on PRESENT. In Proceedings of the 2016 8th IEEE International Conference on Communication Software and Networks, Beijing, China, 4–6 June 2016; pp. 8–12.