Hindawi
Security and Communication Networks
Volume 2022, Article ID 5834419, 12 pages
https://doi.org/10.1155/2022/5834419

Research Article

Remote Surveillance System without Privacy Leakage for Contacts in Infectious Disease

Zechen Li,[1] **Jingxiu Zhao**,[1] **Jing Meng**,[1] **Guangtao Si**,[1] **and Wei Li**[2]

[1]*School of Computer Science, Qufu Normal University, Rizhao 276826, China*
[2]*Department of Computer Science, Georgia State University, Atlanta 30302, GA, USA*

Correspondence should be addressed to Jing Meng; jingmeng@qfnu.edu.cn

Received 16 January 2022; Revised 30 March 2022; Accepted 28 June 2022; Published 22 August 2022

Academic Editor: Jinguang Han

Remote surveillance is an effective method in restraining the spread of infectious diseases via monitoring crowds in affected areas. However, the monitoring targets in existing works are crowds, leading to high system cost, and most of them focus on finding close contacts, less considering the privacy protection and suspected treatment issues. To conquer above problems, this paper develops a new remote surveillance system for infectious diseases, which contains the following major contributions: (1) the monitoring targets are the ordinary contacts not all crowds, effectively decreasing the system cost; (2) establish the joint-control mechanism among contacts, health center, and the hospital to facilitate the prediagnosis and in-time treatment for suspected patients; (3) to avoid the privacy leakage of contacts, a double encryption strategy is designed to protect the location information, and a separating database-storage mechanism is used to improve the contact's data security on the whole. Theoretical security analysis showed that the proposed system and privacy protection methods can effectively fight transmission attacks and avoid privacy leakage during data usage. Based on the created COVID-19 dataset, the simulation experiments were carried out to evaluate the effectiveness and feasibility of the proposed system, on the metrics of the accuracy of prediagnosis, encryption, and decryption time for health and location data. In summary, this work provides a promising way of low-cost remote surveillance system without privacy leakage to control the spreading of infectious disease.

1. Introduction

The outbreak of infectious diseases will cause health, economic crises, and social panic and pose great threats to the stability of society. For example, SARS outbreak in 2003 caused more than 8,000 people in more than 20 countries to be infected; the new coronavirus in 2019 is spreading quickly; it has seriously affected the normal lives of people all over the world and may have deep impact for our future life. In order to suppress the rapid spread of the virus, while providing maximum personal freedom and job security, remote monitoring strategy based on the network communication has emerged. It can collect health information and monitor dynamic trajectories of crowds, thereby effectively inhibiting the spread of the virus.

In the field of medical monitoring, the surveillance system generally collects residents' health data and records their moving trajectory through body sensors [1]. Maglogiannis et al. first proposed to acquire the patients' health information by the sensors equipped on the body and obtain their locations through various indoor and outdoor positioning and tracking technologies [2]. Since then, many researchers have developed various remote monitoring methods to acquire all kinds of information from the patient and applied them in different scenarios [3–6]. Rajavel et al. proposed an IoT-based smart healthcare video surveillance system; it applied edge computing to the abnormal falling activity monitoring for remote patient and elderly people. The using of edge computing reduces the network bandwidth and response time and maximizes the fall behavior prediction accuracy significantly [7]. For diabetic retinopathy, a cloud-based diabetic retinopathy prediction system is proposed using the recurrent convolution neural network classifier model. It can provide risk stratification, optimized

resource allocation, severe disease prediction, and risk control during the screening and diagnosis process [8].

In recent years, the sensor-based remote surveillance technology has been applied to the field of infectious diseases to find suspected patients (also called infectious source) through monitoring the health status and interaction information of related people. For example, Zhang et al. proposed integrating wireless body area networks (WBANs) with mobile phones to collect the body vital signs and use genetic searching and dominating set identification algorithms to effectively identify epidemic sources [9]. Later, this research group divided the populations into multiple clusters based on their physical location and property similarity and then applied different identification algorithm to find the source of infection [10]. In addition, Sanjay et al. developed a surveillance system to monitor the public. In this work, a dynamic social network map to determine the Ebola infected persons was constructed by the collected social interaction information and physical vital signs of people [11]. But in these works, the data privacy issues are not considered. In fact, there exists a lot of sensitive information of users in the health and location data collected by the sensing device, and the leakage of information can easily lead to various criminal activities. Recently, Sandeep et al. employed IoT and fog healthcare systems to identify infected person and control CHV outbreaks, and at the same time, they realized the privacy protection of the user's sensitive information, health data, and location data at different privacy protection levels based on information fragmentation and key information protection mechanism [12].

Although a few works have been developed to find the infectious source based on the surveillance system, the supervision targets in these works are all crowds in the related area. When the spreading of infectious disease covers a large region in the world, such as the coronavirus out breaking in 2019, the comprehensive surveillance will lead to high system cost and inconvenience to many people. Considering that the outbreak of infectious disease is always found by some confirmed patients, and thus we can find the contacts of these persons by their social-interaction history, we proposed a remote surveillance system with a new concept of just monitoring the contacts, leading to low system costs and low influence to most of people. Considering the fact that the risk of infection is different according to the degree of contact, we divide the contacts into ordinary contacts and close contacts based on the distance between them and the confirmed patients. Here, we call people within 1.5 meters' contact distance as close contacts, and those who appear in a risk place but are more than 1.5 meters away are called ordinary contacts. Close contacts will be isolated directly, and others will be under the system monitoring with personal freedom. Different from the existing surveillance system, our proposed system implements the joint cooperation and control among the contacts, health center, and the hospitals, implementing both the data collection of health and location information, the prediagnosis of contacts' health status, and the privacy protection of contacts' sensitive data simultaneously. Theory analysis and simulated experiments verified the feasibility and effectiveness of our proposed surveillance system.

The rest of the paper is organized as follows: Section 2 describes the related work; Section 3 describes the system architecture and related privacy-threat model; Section 4 proposes schemes for data privacy protection and strategies of algorithm designing; Section 5 is security analysis of the proposed system and the designed algorithms; Section 6 conducts experiments to evaluate the data classification methods and system running time; the last section is conclusion and discussions.

2. Related Works

The goal of our work is to monitor the contacts, but finding contacts, i.e., contact tracing, should be performed before our work. Contact tracing is essentially a process of identifying persons who may have come into contact with an infected person; some related works have been reported [13]. In 2016, Qathrady et al. proposed using backward search techniques for tracing of at-risk population [14]. In 2018, Altuwaiyan et al. proposed a solution to find contacts in infectious disease, in which the user's location was recorded by short-range wireless devices; using homomorphic encryption to match common wireless devices between the infected user and regular user, the contacts can be determined by weighted score matching method [15]. To control the spread of the disease and avoid the second wave of infection, Paulo Klaine et al. proposed a new blockchain-based contact tracking framework [16]. Recently, Zhang et al. proposed a contact tracing scheme in 5G-integrated and blockchain-based medical applications, in which public can perform location checking with their mobile phones or even wearable devices connected to 5G network to find whether they have been in possible contact with a diagnosed patient [17]. In addition, researchers use sensor monitoring to find susceptible people; for example, Zhang et al. proposed an analysis method of interpersonal infection, which realized the judgment of susceptible people by analyzing the monitored data (physical condition, length of contact, distance, etc.) [18]. Finding the contacts is out of the goal of our work, but it is indeed the preconducted task for the monitoring of contacts, and the existing methods can be easily integrated into our proposed system.

Some privacy-preserving schemes in remote monitoring system for infectious disease have also been developed. Liu et al. used a key-independent inner product encryption mechanism to ensure that untrusted entities can only obtain health statistics but not personal data, thereby realizing the protection of user's health and location data [19]. However, this work only achieved statistical analysis of health data, and the user's real location information was not obtained, so it cannot find new contacts. Zhang et al. implemented the protection on users' health and social network data by homomorphic encryption while finding the contacts [18], but this work did not consider the issues of users' location privacy. In [16], due to user's temporary pseudo-ID periodically generated at random, it prevents the user location from being tracked, preserving users' location privacy.

Altuwaiyan et al. [15] also proposed a solution to find contacts in infectious disease, in which the user's location was protected by homomorphic encryption method, but this work is not involving health data monitoring. Aiming at the above issues, Sandeep et al. proposed a comprehensive data collection and privacy protection scheme for Chikungunya disease (CHV). It employed IoT and fog healthcare systems to identify and control CHV outbreaks and used information fragmentation and key information protection mechanism to realize the privacy protection of the user's sensitive information, health data, and location data at different privacy protection levels [12]. However, the goal of this system is still to monitor all residents in the affected area.

Compared with the previous work, this paper further classifies the contacts of infectious disease into two cases of close contacts and ordinary contacts and designs a complete remote monitoring system model for ordinary contacts. It simultaneously fulfills the requirements of health data monitoring, location trajectory tracking, data privacy protection, and suspected treatment. To guarantee the security of joint performance among contacts, health center, and hospitals in our proposed system, the privacy protection strategy for contacts' sensitive information in transmission and storage was developed. Specifically, an asymmetric encryption method was employed to protect the contacts' health data, and a double encryption mechanism and separating database storage were presented to implement the privacy protection of location information. In the experiment, a synthetic COVID-19 dataset was constructed by ourselves to verify the performance of the proposed system.

The system model proposed in this paper takes the health center as the core unit of data storage and analysis and is responsible for the communication between the monitored person and the hospital. It can be applied directly for the scenario where infectious disease outbreaks in a relative small area. Moreover, for the case of a national or even global epidemic of infectious disease, our proposed system can also be a component in the whole monitoring system, is responsible for the surveillance for a local area, and then sends the valuable data of this area to higher-level server that stores aggregate data.

3. System Model

3.1. System Architecture. This paper designs a remote monitoring system for ordinary contact in infectious disease, as shown in Figure 1; it simultaneously achieves the collection of contacts' health information, monitoring of their moving trajectory, and in-time treatment of suspected patients.

This model includes four entities: contacts, health centers, hospitals, and trusted authority. Contacts, health center, and hospitals firstly register at trusted authority and obtain their legal identity ID. In the performance of the system, contacts collect their own health data and location information through various wearing sensors and send them to the health center. Then, the health center will receive and store the health and location data and further implement the preliminary diagnosis on the contacts based on these data. Once a contact is diagnosed as suspected one, the health center will match the nearest hospital for the suspected contacts, and the hospital is responsible for the treatment of the suspected patients.

Major steps in this system are summarized as follows:

(1) Registration. Contacts, health center, and legal hospital register at the trusted authority (TA), respectively. TA will generate a pseudo-identity ID and a burry home address of the contact and share this information and contact information with the health center. Moreover, trusted authority will generate valid keys for legal entities to encrypt and decrypt the data.

(2) Collection of health and location data. The contacts collect their health and location data by equipped sensors at regular intervals and then send them to the health center.

(3) Data storage and prediagnosis. The health center stores the health and location data from contacts into the health and location databases, and it also makes prediagnosis to find the suspected patients.

(4) Hospital allocation. Once a contact is diagnosed as a suspected patient, the health center will assign the nearest hospital to the contact and send the contact information and health information of the suspected contact to the hospital at the same time.

(5) Treatment of suspected patients. After the hospital receives the information of a suspected patient, it sends a request of precise location with its ID to the patient, and after the suspected patients verify the legality of the allocated hospital, he will send his exact location to the hospital.

(6) Extraction of moving trajectory. Once the suspected patient is confirmed to be infected, the health center will extract the moving trajectory of the patient in the location database to search related contacts. Here, how to find new contacts based on the trajectory data is not the task of our work.

3.2. Threat Model. Assuming the hospital is completely reliable and will not disclose the patient's private data, two kinds of privacy-leakage threats in the proposed monitoring system are indicated in the dashed-line rectangles in Figure 1.

(1) The health center is considered semicredible. Staffs honestly obey the confidentiality agreement, but the internal staffs may curiously view and further infer about the data submitted by contacts, leading to the leakage of contacts' sensitive information, such as home address and moving trajectory.

(2) Eavesdropping attacks may occur during data transmission in the network.

In view of the above attack model in the proposed system, our privacy issues should cover the contact's personal information, health, and location data, while ensuring

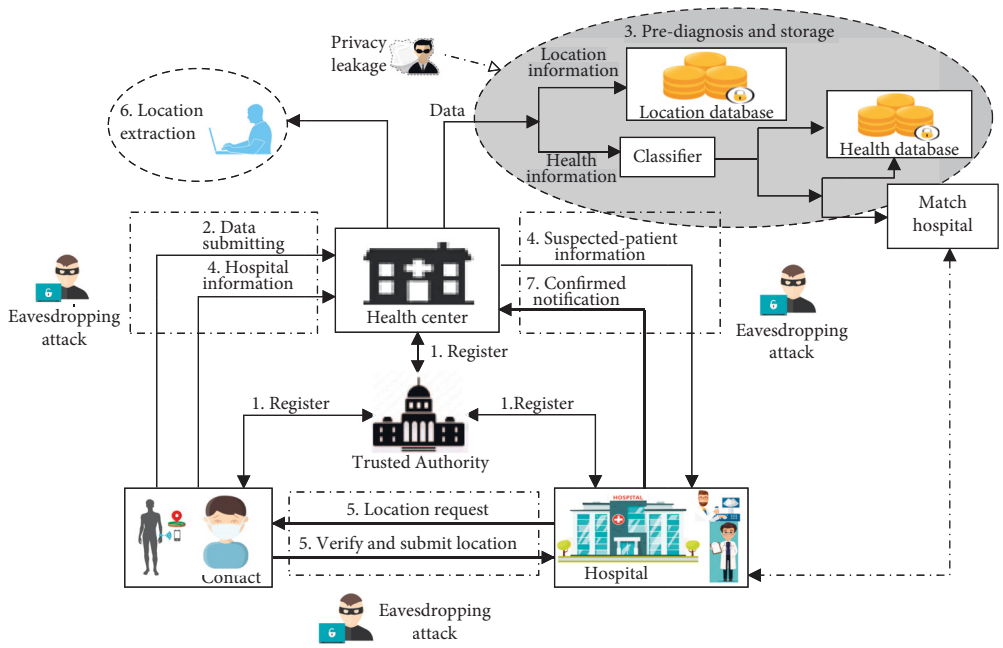

Figure 1: Architecture of the remote surveillance system on contacts.

the reliability and usability of these data. The specific privacy protection goals are as follows:

(1) Confidentiality. The proposed system should ensure that all sensitive information of contacts will not be disclosed to any unauthorized users or potential attackers.

(2) Reliability. To prevent contacts from providing false information to health center, the monitoring system must verify the users' identity and guarantee the correctness of the submitted health data and location information.

(3) Availability. Data availability must be guaranteed while implementing the data protection. For example, the legal member of health center can obtain the real moving trajectory of the contacts, while the location data are protected from being attacked by others.

4. Privacy Protection Method

In this section, we introduce the privacy protection scheme for health and location data produced in our proposed system. To better explain the methods, all notations and functions used in this work were listed in Table 1.

4.1. Personal Information. In our proposed system, the contact will firstly register at the fully trusted authority (TA) using his personal information; then the TA returns a pseudo-ID to the contact and shares the information, contact's contact information, and his burry address with health center at the same time for subsequent operations. Therefore, only the TA requires the real personal information of the contacts. Due to the trusty of TA, the contact's personal information is security. But to avoid the privacy-leakage of personal data during transmission, the function E_{reg} is used to encrypt the registration data and submit it to the trusted authority. Once the monitored contact becomes a suspected patient, the allocated hospital sends a location request to the contact, and the contact will use the function $E_{a\ dd}$ to encrypt his home address and send it to this hospital. Above process is shown by the solid line in Figure 2.

Table 1: Notations and functions.

Notations	Functions
PK_{hc}	Health center's public key
PK_h	Hospital's public key
SK_{hc}	Health center's private key
k_{sym}	User's session key
Ba	Contact's blurry home address
C	Contact's contact information
HI	Contact's health data
LI	Contact's location data
PI	Contact's personal information
N_CD	Notification of confirmed diagnosis
E_{reg}	Register function
$E_{a\ dd}$	Address transfer function
E_{hi}	Health data upload function
E_{ci}	Data-transfer function of suspected user
E_k	Transfer function of session key
E_{li}	Double encryption for location data
E_{asym}	First-layer asymmetric encryption for LI
E_{sym}	Second-layer symmetric encryption for LI
D	Decryption function for LI

4.2. Health Data. In order to prevent unauthorized users from prying into health data stored in the health center and avoid the risk of eavesdropping during data transmission,

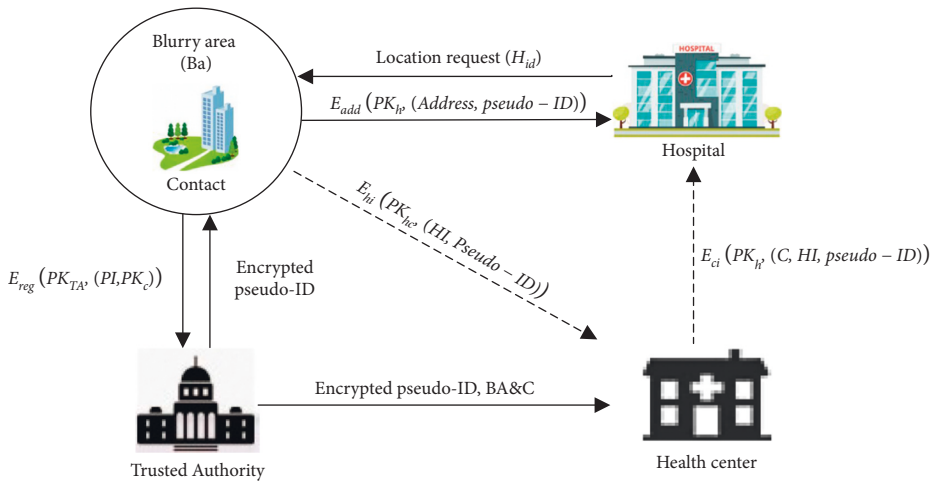

Figure 2: Flowchart of privacy protection on personal and health data.

this paper employed an asymmetric encryption method of elliptic curves cryptography (ECC) to encrypt the health data. First, the contact's pseudo-ID and the public key of the health center are written into the wireless sensors equipped on the contact. Then, health data are automatically collected at regular intervals and encrypted by ECC. Finally, these data are uploaded to the health center by function E_{hi} as shown in Figure 2. After receiving the health information, the staff in the health center can decrypt them with the private key SK_{hc} to obtain the real health data for prediagnosis analysis. When the contact is determined to be a suspected one, his health data will be encrypted and sent to the allocated hospitals, which is implemented by function E_{ci} in Figure 2.

The pseudocodes implementing the privacy protection on personal information and health data were given in Algorithm 1; it involves 4 subfunctions: user registration function E_{reg}, health data upload function E_{hi}, data-transfer function E_{ci}, and address transfer function $E_{a\ dd}$. The pseudocodes of these 4 functions are given in Algorithms 2 to 5.

Algorithm 2 describes the user registration process, in which Tel is the contact's telephone number, Ba represents the blurry area of address, and PK_c is the contact's public key.

Algorithm 3 describes the procedure of uploading health information. E_{hi} is used to encrypt the health data collected at different time i.

When a contact is diagnosed as a suspected patient, the health center transmits his relevant data to the hospital for further diagnosis, as illustrated in Algorithm 4

Algorithm 5 describes the address transfer process between the contact and the hospital. After receiving the location request, the diagnosed person verifies the identity of the hospital by comparing the $H_{i\ d}$ sent by the hospital with the $H_{i\ d}{}'$ provided by the health center. If the identity is legal, the specific home location is encrypted and sent to this hospital.

4.3. Location Information. In the process of remote monitoring, the contact's location information will be collected at regular intervals and stored in the health center. Because the accessing on location data only occurred when the contact is confirmed, the health data may be visited many times (such as for medical study). In addition, the leakage of location information may expose other sensitive information (such as inferring the contact's address and real identity), and further resulting in the risk of health data with the connection of pseudo-ID. So, a separate storing strategy was designed; i.e., the location data are stored in another database, which can improve the security of the contact's information on the whole.

To ensure the security of the users' location information, we designed a double encryption mechanism combining asymmetric encryption and symmetric encryption in this work, which is implemented by function E_{li}. First, when contacts upload their location information to the health center, in order to prevent unauthorized persons in the health center from prying on the users' location information, a first-layer symmetrical encryption mechanism is designed; i.e., original location data is encrypted using the session key implemented by E_{sym}. Second, the second-layer asymmetrical encryption is performed on the encrypted location information by E_{asym}, to protect its security in data transmission. In this case, unauthorized persons in the health center cannot decrypt the contact's location information because they do not have the session key. On the other hand, when a contact is confirmed, the contact will encrypt the session key by function E_k and send it to the authorized person; the health center will receive a notification of confirmed diagnosis (N_CD) from the hospital at the same time. Then, the authorized ones of the health center can extract the patient's location data with the session key from the user, which is implemented by function D. The flowchart of above process is shown in Figure 3.

Algorithm 6 lists the pseudocodes of the designed algorithm in Figure 3. It includes two subfunctions to implement the encryption and decryption process of users' location data. The pseudocode for these two functions is given in Algorithms 7 and 8.

Algorithm 7 describes the location information encryption upload process. For location information collected at different times i, the location information is encrypted

(1) user registration by $E_{reg}(PK_{TA}, (PI, PK_c))$
(2) if user is legal
(3) obtain the pseudo-ID from authority
(4) else
(5) reject the users' registration request and exit
(6) write the pseudo-ID and PK_{hc} into wireless sensor devices
(7) for each health data at timestamp
(8) data encryption by $E_{hi}(PK_{hc}, (HI, pseu\ do - I\ D))$
(9) data storage and pre-diagnosis analysis in health center
(10) if HI is abnormal
(11) send the related information to hospital by $E_{ci}(PK_h, (C, HI, pseu\ do - I\ D))$
(12) hospital sends the location request to the contact with its $H_{i\ d}$
(13) contact verifies the legality of the hospital by its identity ID
(14) if $H_{i\ d}$ is legal
(15) contact sends address to hospital by $E_{a\ dd}(PK_h, (A\ dd\ ress, pseu\ do - I\ D))$
(16) else
(17) contact rejects the location request
(18) end if
(19) end if
(20) end for

Algorithm 1: Personal information and health data protection.

Input: Tel, Ba, PK_C
Output: flag (success 1, otherwise 0)
(1) flag = 0
(2) $RI \leftarrow (Tel, Ba, PK_c)$
(3) encrypted RI with PK_{hc}
(4) flag = register in health center using RI
(5) If flag
(6) obtain Pseudo-ID from the health center
(7) return flag
(8) end if

Algorithm 2: User registration.

Input: PK_{hc}, pseudo-ID, HI
Output: flag (success 1; otherwise 0)
(1) E_{hi} = encrypted HI with PK_{hc}
(2) flag = submit E_{hi} to health center with Pseudo-ID
(3) data analysis and storage
(4) return flag

Algorithm 3: Health information uploading.

Input: PK_h, Tel, HI
Output: flag (success 1; otherwise 0)
(1) E_{hi} = encrypted Tel and HI with PK_h
(2) flag = submit E_{ci} to hospital
(3) return flag

Algorithm 4: Deliver data to hospital.

Input: PK_h, address
Output: flag (success 1, otherwise 0)
(1) get $H_{i\,d}{}'$ from health center
(2) verify the hospital ID by comparing $H_{i\,d}{}'$ and $H_{i\,d}$
(3) if legal
(4) $E_{a\,dd}$ = encrypted home address with PK_h
(5) flag = send $E_{a\,dd}$ to hospital
(6) end if
(7) return flag

Algorithm 5: Address transfer between contact and hospital.

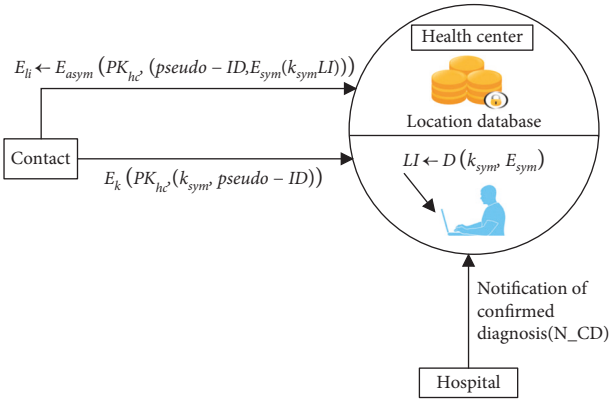

Figure 3: Flowchart of privacy protection on location information.

(1) for each location data LI collected at each timestamp
(2) contact encrypts the location data by $E_{li} = E_{asym}(PK_{hc}, (pseu\ do - I\ D, E_{sym}(k_{sym}, LI)))$
(3) location database $\leftarrow E_{li}$
(4) end for
(5) if contact is confirmed
(6) hospital sends N_CD of the contact to the health center
(7) user encrypts the k_{sym} by $E_k(PK_{hc}, (k_{sym}, pseu\ do - I\ D))$, and sends it to health center
(8) health center decrypts the location data of the contact by $LI \leftarrow D(k_{sym}, E_{sym})$
(9) end if

Algorithm 6: Location information protection.

Input: k_{sym}, PK_{HC}, LI
Output: flag (success 1; otherwise 0)
(1) location information acquisition at time i
(2) for each LI_i acquired at time i
(3) $E_{li_{l1}} = E_{asym}(PK_{hc}, LI_i)$
(4) $E_{li_{l2}} = E_{sym}(k_{sym}, E_{li_{l1}})$
(5) flag = submit $E_{li_{l2}}$ to health center
(6) return flag
(7) end for

Algorithm 7: Location information uploading.

Input: E_{li}, SK_{HC}, k_{sym}
Output: T_{li} (the total location information)
(1) if contact is confirmed
(2) $E_k(PK_{hc}, k_{sym})$, then send it to health center
(3) for each E_{li_i} at time i of the contact
(4) $E_{li_{l1}} = D_1(k_{sym}, E_{li_i})$
(5) $LI = D_2(SK_{hc}, E_{li_{l1}})$
(6) $T_{li} = T_{li} \cup LI$
(7) end for
(8) return T_{li}
(9) end if

Algorithm 8: Location information decryption.

using function E_{asym} for the first layer, and function E_{sym} is used for the second encryption and then uploaded to the health center for storage.

Algorithm 8 describes the location information decryption process. When the contact is diagnosed, it sends k_{sym} to the authorized personnel of the health center. The health center uses the D_1 function to decrypt the first layer and then uses the D_2 function to decrypt the second time and finally get the total location information LI.

5. Security Analysis

5.1. Data Security in Health Center. The system framework proposed in this work can effectively realize the privacy protection of users' data stored in the health center. First, contact registers at the TA using his personal information, so his personal identity can be protected due to the trusty of TA. We assume that the staffs in health center have no way to retrieve other information of user through his contact information, so based on the proposed user registration mechanism, health center can only obtain the pseudo-ID and contact information, not knowing the true user identity of the health data. So when the employees in health center pry the health data, they cannot infer the user's any other sensitive information. Second, a double encryption mechanism was designed to protect the users' location data. In this strategy, only the authorized persons in health center can obtain the user's session key and extract the user's moving trajectory. It can be seen that this method effectively avoids the curious prying and identity inference of the staff in health center through the user's trajectory information. Third, the proposed system uses wireless sensor equipment to automatically collect, encrypt, and upload medical data to health center at regular intervals, which effectively prevents users from tampering with the data. Finally, we also implemented the identity verification on users and hospitals; i.e., when the health center receives any data from the contact or hospital, it first verifies his identity by pseudo-ID, and the service is provided only when the identity is verified.

5.2. Data Security during Transmission. To prevent data eavesdropping during data transmission, this work employed data encryption mechanism. First, all data uploaded to the health center are encrypted with the public key of health center. Because other people do not have the private key of health center, even if the data are eavesdropped during transmission, they cannot decrypt them. Second, when contacts or health centers transmit data to the hospital, they use the hospital's public key to encrypt the data before transmission; when the hospital needs to send the diagnosis result to the health center, it uses the public key of the health center to encrypt the data before transmission. Therefore, other persons cannot decrypt the real information without the hospital's private key, even if they obtain the data. It can be seen that the encryption mechanism used in this article effectively guarantees the security of the data during transmission.

6. Experiments and Results

In Section 5, the security of the proposed system and methods on protecting the users' sensitivity information have been discussed. In this section, we take COVID-19 as an example to evaluate the performance of our proposed system. Because we did not find any labeled dataset on COVID-19 patients with related symptoms, we created labeled dataset by some data collected in network. Then, using the labeled health data of COVID-19 and some simulated location data, we discussed four classification methods for prediagnosis and evaluated the running time of several key steps in our surveillance system. All experiments were carried out on a PC (Intel i5-6200U CPU, 2.30 GHz, 8G memory) with python 3.7. The experiments include three parts: creation of synthetic data set of COVID-19, discussion on classification algorithms, and evaluation on the running time of system.

6.1. Data Creation and Prediagnosis. Taking COVID-19 as an example, we conducted a relatively comprehensive search on the Internet and found six symptoms related to the COVID-19 and their correlation coefficients reflecting the incidence probability of this disease based on the clinical statistics, as shown in Table 2 [20]. We did not find any labeled dataset on COVID-19 patients with above symptoms, so we created labeled dataset by synthetic data including the symptom information and patients' other

Table 2: Symptoms and related incidence probability for COVID-19.

Symptoms	Probability
Fever	0.661
Tiredness	0.421
Dry cough	0.579
Difficulty in breathing	0.363
Sore throat	0.24
None symptom	0.058
Pains	0.427
Nasal congestion	0.041
Runny nose	0.094
Diarrhea	0.088
Headache	0.374

information. The flowchart of dataset creation is given in Figure 4, and the major steps can be summarized as follows.

Step 1: Compute symptom score by summing the correlation coefficients of existing symptoms of the sample.

Step 2: Considering some other factors (age, history of COVID-19 disease, and history of exposure) also have important influences on the incidence probability of COVID-19, we compute a global score by multiplying the symptom score and weights of these factors. Here, the values of weights are set by ourselves based on the related reports, as shown in Table 3.

Step 3: Normalize the global score into the range between 0 and 1, so that it can reflect the incidence probability.

Step 4: Set a threshold on global score to classify the data into two kinds: normal and suspected ones.

Step 5: Combine the synthetic data into the labeled dataset, if it is a new instance.

To implement the prediagnosis of the contact, four classification methods, including k-nearest neighbor, support vector machine, decision tree, and random forest, were employed to classify the contacts' health data and provide the result of prediagnosis. In our experiments, the synthetic dataset consists of 4225 cases; each case involves 11 basic symptoms related with COVID-19 and 3 other personal attributes. Specifically, 80% synthetic data were used to optimize the classifiers, and the remaining 20% data were used to evaluate their performance. The accuracy on classification of the four methods using different threshold on global score was listed in Table 4. It can be seen that all the four classification methods can provide satisfactory prediagnosis accuracy, and the random forest method achieves the best. So, we choose this method to implement the prediagnosis based on the health data in our proposed monitoring system.

6.2. Evaluation on Running Time. The running time of our proposed remote surveillance system is composed of two parts: one is the data transmission among contact, health center, and hospital; the other is the encryption computation on various data. The data transmission time is controlled by the communication network, and thus we focus on evaluating the data-processing time of each entity.

For contact, we mainly evaluate the encryption time for personal information in registration process, health information, and location information on daily uploading. In our experiments, ECC method was used to realize the encryption on the contact's personal information and health data; a double encryption method combining ECC and advanced encryption standard (AES) was developed to realize the encryption of location data. In the registration process, including the contact's personal information, the size of these data is about 26 bytes, and the encryption time is about 0.000964 s. The health information of the contact is acquired and submitted to the health center at predefined intervals (one day in our experiments). When COVID-19 health data is used as an example, it contains 14 symptoms' data related to this disease with data size of about 178 bytes, and the evaluated ECC encryption time is about 0.001001 s. Once the health information of the contact is abnormal, the health center packs the health data of the contact in recent days (14 days in our experiment) to the hospital. In this case, the total amount of health data is about 2505 bytes, and the ECC encryption time is evaluated about 0.002982 s, just increased about two times of the case of one day, and did not affect the practicality of the surveillance system. To track the moving trajectory of the contact, his location information should be acquired at a predefined interval; it was represented as a vector (longitude, latitude) with size of 50 bytes. The double encryption time on single-time location data is evaluated about 0.003963 s. It can be seen that although the encryption methods are usually time-consuming, the time spent on the data encryption in this work is small and can satisfy the requirements of system, for it was conducted just on single-time small-amount health and location data.

For health center, we mainly evaluate the decryption time for health data and location information and the prediagnosis time of the health information. The health center decrypts the health information sent by the contacts every day; the ECC decryption time is evaluated about 0.000998 s. When the suspected patient is confirmed to be infected, the health center will extract and decrypt the moving trajectory of the patient, the amount of location information that needs to be decrypted depends on the specific situation, we assume that the user's location information is recorded every two hours from 6 am to 10 pm in our experiment, and we evaluated the decryption time on single-time location data to be about 0.002031 s, so the decryption time on 14 days of location information is about 0.227472 s. Here, the random forest method is chosen to implement the prediagnosis of the contact, and its classification time is about 0.4 seconds. All running time on different data from contact and health center is listed in Table 5. It can be seen that the encryption and decryption time for data from single-time point is short for the health data and location information. Although the time spent on the data collected in the 14 days is dramatically longer than that from one time point, such that the encryption time of

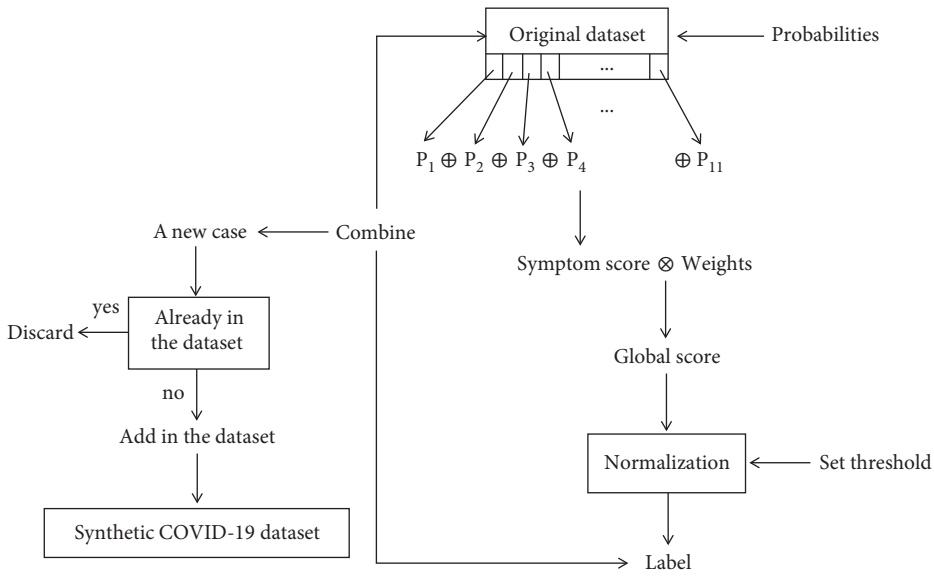

Figure 4: Flowchart of creating synthetic COVID-19 dataset.

Table 3: Other factors and weights.

Age		History of COVID-19		History of exposure	
Low-risk-age	1.3	Yes	0.8	Contact-yes	2
Moderate-risk-age	1.6	No	1.2	Contact-no	1
Sever-risk-age	1.9			Contact-unknown	1

Table 4: Prediagnosis accuracy of different classification methods.

Thresholds	Confirmed cases	Normal cases	*K*-nearest neighbor (%)	Support vector machine (%)	Decision tree (%)	Random forest (%)
0.20	1390	2834	92.83	92.25	92.18	95.12
0.25	966	3258	94.55	95.55	91.54	95.55
0.30	698	3526	95.05	96.13	92.11	96.92
0.35	487	3737	96.13	96.84	94.84	96.77

Table 5: Evaluation on running time for different entities.

Entity	Process step	Data content	Execution time (s)
Contact	Encryption	Personal information	0.000964
		Health data (one time)	0.001001
		Location information (one time)	0.003963
Health center	Encryption	Health data (14 days)	0.002982
	Decryption	Health data (one time)	0.000998
		Location information (one time)	0.002031
		Location information (14 days)	0.227472
	Prediagnosis		0.453125

health data for 14 days is approximately three times the one time, the decryption time of location data for 14 days is approximately 100 times the one time. But it is still far less than one second. As a result, the time spent on the data encryption and decryption in this work is small and can satisfy the requirements of system performance.

Moreover, compared with other remote surveillance systems on crowds, our proposed system model aiming at the ordinary contacts can dramatically decrease the system loading. Taking COVID-19 as an example: the population of China is about 1.4 billion; according to Internet data, from the outbreak of the infectious disease to its peak time of

August, 2020, the cumulative number of contacts in China is about 800,000, accounting for five in ten thousands of the total population.

7. Discussion and Conclusion

7.1. Discussion. Aiming at the problems of relative small functions and the large amount of monitoring objects in the existing remote surveillance system, a new concept of "ordinary contacts" is defined based on the contact distance between the people and the confirmed patients. Then, a novel remote monitoring system for ordinary contacts is proposed, leading to reduced system costs. This work not only realizes the monitoring and analysis of users' health status and location trajectory, but also protects the users' privacy of their sensitive data. Moreover, it implements the in-time treatment on suspicious patients via the cooperation among health center, hospitals, and contacts.

To better understand our work, three issues were discussed as follows.

(1) The kernel tasks of our work are the tracking, monitoring, and privacy protection on the data of ordinary contacts. Finding new contacts of a confirmed patient via his moving trajectory is out of the scope of our work, and the related methods can be found in [14, 15, 18].

(2) In this paper, we mainly adopt data encryption technology to implement the users' privacy protection, which can effectively ensure the security of data in the process of transmission, storage, and usage, but it will be time-consuming for large-size data. Fortunately, the encryption operation was used for the text data of single-time acquisition in our proposed system, and in this case, the data amount is small and thus the encryption time is short, meeting the requirements of the practical applications. However, with the increase of the data amount, for example, health data containing images or videos, the encryption time will increase and the response speed of the system will decrease at the same time.

(3) In our work, the local server is used to store the text data collected in the surveillance system; the storage capacity meets the practical requirements for monitoring on local area. For the health monitoring on large region with more data types (such as images) collected and analyzed, the cloud-based server may be required.

7.2. Conclusion. In summary, this paper proposes a novel remote monitoring system for ordinary contacts of infectious diseases, which not only monitors and analyzes users' health status and location trajectories, but also protects the privacy of users' sensitive data. In addition, prompt treatment of suspected patients is implemented through cooperation among health centers, hospitals, and contacts. Compared with the existing work, this work can effectively decrease the system cost by just monitoring the ordinary contacts and simultaneously implements the data protection, prediagnosis, and treatment of contacts. To further improve the performance of the system in various scenarios, several potential future directions are given as follows: (1) In the future, we will explore new privacy protection strategies adaptive to large-size data, such as images; (2) to expand the application scope of the system, cloud-server based remote surveillance system will be developed; it can be applied for large-area surveillance and large-amount data storage and analysis; (3) currently, our experiments were conducted on some generated dataset, for the difficulty in establishing a real system platform. Next, based on mobile phone software, Bluetooth technique, and joint data server, we will try to build a running environment for the proposed system architecture and further verify the related methods used in the system.

Data Availability

The data used to support the findings of this study are included within the article.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or nonfinancial interest in the subject matter or materials discussed in this manuscript.

Authors' Contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Zechen Li, Jingxiu Zhao, Jing Meng, Guangtao Si, and Wei Li. The first draft of the manuscript was written by Zechen Li and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work is supported by Natural Science Foundation of Shandong Province (no. ZR2020MF105).

References

[1] S. Patel, H. Park, P. Bonato, L. Chan, and M. M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, p. 21, 2012.

[2] I. Maglogiannis, L. Kazatzopoulos, K. Delakouridis, and S. Hadjiefthymiades, "Enabling location privacy and medical data encryption in patient telemonitoring systems," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 946–954, 2009.

[3] X. Liang, L. Xu, R. Lu, X. Lin, and X. Shen, "Enabling pervasive healthcare with privacy preservation in smart community," in *Proceedings of the 2012 IEEE International Conference on Communications*, pp. 3451–3455, Ottawa, ON, Canada, June 2012.

[4] K. A. A. Mamun, M. Alhussein, K. Sailunaz, and M. S. Islam, "Cloud based framework for Parkinson's disease diagnosis and monitoring system for remote healthcare applications," *Future Generation Computer Systems*, vol. 66, pp. 36–47, 2017.

[5] C. Li, X. Hu, and L. Zhang, "The IoT-based heart disease monitoring system for pervasive healthcare service," *Procedia Computer Science*, vol. 112, pp. 2328–2334, 2017.

[6] N. Jalloul, "Wearable sensors for the monitoring of movement disorders," *Biomedical Journal*, vol. 41, no. 4, pp. 249–253, 2018.

[7] R. Rajavel, S. K. Ravichandran, K. Harimoorthy, P. Nagappan, and K. R. Gobichettipalayam, "IoT-based smart healthcare video surveillance system using edge computing," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 6, pp. 3195–3207, 2021.

[8] G. Sundharamurthy and V. K. Kaliappan, "Cloud-based onboard prediction and diagnosis of diabetic retinopathy," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 24, pp. 1–11, 2021.

[9] Z. Zhang, H. Wang, X. Lin, H. Fang, and D. Xuan, "Effective epidemic control and source tracing through mobile social sensing over WBANs," in *Proceedings of the International Conference on Computer Communications*, pp. 300–304, Turin, Italy, April 2013.

[10] Z. Zhang, H. Wang, C. Wang, and H. Fang, "Cluster-based epidemic control through smartphone-based body area networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 3, pp. 681–690, 2015.

[11] S. Sareen, S. K. Sood, and S. K. Gupta, "IoT-based cloud framework to control Ebola virus outbreak," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 3, pp. 459–476, 2018.

[12] S. K. Sood and I. Mahajan, "Wearable IoT sensor based healthcare system for identifying and controlling chikungunya virus," *Computers in Industry*, vol. 91, pp. 33–44, 2017.

[13] L. Ferretti, C. Wymant, M. Kendall et al., "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Science*, vol. 368, no. 6491, Article ID eabb6936, 2020.

[14] M. A. Qathrady, A. Helmy, and K. Almuzaini, "Infection tracing in smart hospitals," in *Proceedings of the 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1–8, New York, NY, USA, October 2016.

[15] T. Altuwaiyan, M. Hadian, and X. Liang, "Efficient privacy-preserving contact tracing for infection detection," in *Proceedings of the 2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, Kansas City, MO, USA, May 2018.

[16] P. V. Klaine, L. Zhang, B. Zhou, Y. Sun, H. Xu, and M. Imran, "Privacy-preserving contact tracing and public risk assessment using blockchain for COVID-19 pandemic," *IEEE Internet of Things Magazine*, vol. 3, no. 3, pp. 58–63, 2020.

[17] C. Zhang, C. Xu, K. Sharif, and L. Zhu, "Privacy-preserving contact tracing in 5G-integrated and blockchain-based medical applications," *Computer Standards & Interfaces*, vol. 77, no. 14, Article ID 103520, 2021.

[18] K. Zhang, X. Liang, J. Ni, K. Yang, and X. S. Shen, "Exploiting social network to enhance human-to-human infection analysis without privacy leakage," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 607–620, 2018.

[19] J. Liu, Y. Hu, H. Yue, Y. Gong, and Y. Fang, "A Cloud-based secure and privacy-preserving clustering analysis of infectious disease," in *Proceedings of the 2018 IEEE Symposium on Privacy-Aware Computing (PAC)*, pp. 107–116, Washington, DC, USA, September 2018.

[20] A. Sarker, S. Lakamana, W. Hogg-Bremer, A. Xie, M. A. Al-Garadi, and Y.-C. Yang, "Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource," *Journal of the American Medical Informatics Association*, vol. 27, no. 8, pp. 1310–1315, 2020.