This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2016.2626288, IEEE Transactions on Dependable and Secure Computing

Exploiting Social Network to Enhance Human-to-Human Infection Analysis Without Privacy Leakage

Kuan Zhang, *Student Member, IEEE*, Xiaohui Liang, *Member, IEEE*, Jianbing Ni, Kan Yang, and Xuemin (Sherman) Shen *Fellow, IEEE*

Abstract—Human-to-human infection, as a type of fatal public health threats, can rapidly spread in a human population, resulting in a large amount of labor and health cost for treatment, control and prevention. To slow down the spread of infection, social network is envisioned to provide detailed contact statistics to isolate susceptive people who has frequent contacts with infected patients. In this paper, we propose a novel human-to-human infection analysis approach by exploiting social network data and health data that are collected by social network and e-healthcare technologies. We enable the social cloud server and health cloud server to exchange social contact information of infected patients and user's health condition in a privacy-preserving way. Specifically, we propose a privacy-preserving data query method based on conditional oblivious transfer to guarantee that only the authorized entities can query users' social data and the social cloud server cannot infer anything during the query. In addition, we propose a privacy-preserving classification-based infection analysis method that can be performed by untrusted cloud servers without accessing the users' health data. The performance evaluation shows that the proposed approach achieves higher infection analysis accuracy with the acceptable computational overhead.

Index Terms—Social network, infection analysis, privacy preservation

1 Introduction

Infectious diseases, such as swine flu, Ebola, and acute respiratory infection, can rapidly spread from human to human within a short period. In 2013, over $200,000$ Canadians get infected with these highly contagious diseases annually, while more than $8,000$ infected patients die as a result [1]. The outbreaks of these infectious diseases usually occur when the infected patients cough and sneeze around non-infected people [2], [3]. It is observed that people with strong social-ties and having frequent or long-lasting contacts (e.g., students studying in the same classroom, and families living in the same house) is likely to spread infectious diseases from the biomedical and sociology perspective [4]. One traditional approach to prevent the spread of infectious diseases [1] is to isolate susceptible patients (who travel from the infected region or have frequent contacts with infected patients during the outbreak season) from the public for a certain period (e.g., two or three weeks) depending on the latent time period of the diseases.

However, this traditional infection prevention approach incurs a large amount of the governmental health expense and labor costs, the isolated patient's economic loss, and even anxiety or panic of the society. To resolve this type of public health crisis, a promising wearable cyber physical system [5] associated with e-healthcare technologies emerges to continuously monitor users' real-time health parameters (temperature, heart rate, electrocardiogram (ECG), etc.), which are formatted in image, audio and text. These health data are collected by a server to analyze abnormal phenomena and provide supporting information for doctor's diagnosis [6]. Although such a wearable cyber physical system is helpful to analyze user's health condition, i.e., whether a user is already infected or not, it lacks sufficient social information to infer the spread of infectious diseases, i.e., whether the user has a high probability to get infected from others.

Social network can offer various applications to mine users' social data during their social interactions [7], [8]. For example, the built-in face-tagging function of Facebook application can identify user's face in pictures and infer if certain users have close social relationships; Wechat friend discovery program can find users in the physical proximity and record social interactions; speech recognition can help to detect if some people cough or sneeze. The fusion of these social network data associated with the monitored health data can provide a novel paradigm to enhance infection analysis. Suppose a junior school student Bob is continuously monitored from both health and social perspectives during the outbreak of infectious disease. Once Bob's immunity strength goes very low and he frequently contacts an infected student, he may be inferred as a susceptible patient in the early stage. The health and social multimedia data are usually collected and processed by multiple independent service providers, such as health institution and social networking service provider (e.g., Facebook and Wechat), respectively. The collaboration of these service providers becomes essential to enable data sharing and processing [9], especially when the volume of continuously monitored data keeps increasing. Incorporating health cloud server collecting users' health parameters and social cloud server, which is maintained by social network service provider to collect users' social networking data including social contact and relationships between users, we envision that the infection analysis can be enhanced.

- *Kuan Zhang, Jianbing Ni, Kan Yang and Xuemin (Sherman) Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada (e-mail: {k52zhang, j25ni, k62yang, sshen}@uwaterloo.ca).*
- *Xiaohui Liang is with the Department of Computer Science, University of Massachusetts at Boston, Boston, USA (e-mail: Xiaohui.Liang@umb.edu).*

Meanwhile, users' health and social data, such as infection status and social contact, are privacy-sensitive [10], [11], and many users are not willing to excessively reveal this private information to the untrusted or unauthorized entities [12]. If the health data and social data are sent to cloud servers in clear text, the untrusted cloud servers may track users' health condition, identity, profiles, contact and social activities, resulting in severe privacy violations, especially for the infected or susceptible patients during the outbreak of infectious diseases. To preserve data privacy, users could encrypt their data and send the ciphertexts to cloud servers [13]. However, this approach may limit the data processing capability of cloud servers [14] and even disable the infection analysis. Therefore, it is challenging to enable the infection analysis and preserve user's privacy at the same time. In addition, social networking data contain some sensitive information of infected and susceptible patients, such as identity and contact details, which may be inferred by the social cloud server when these data are shared to other entities for further health analysis. For example, if the hospital or public health agency queries an infected patient's data on the social cloud server, the social cloud server may infer that the queried user is infected. Meanwhile, the hospital without the user's authorization should not be able to query non-infected user's social networking data. Without sufficient privacy protections, users may not want to share their social and health data to the untrusted cloud servers for infection analysis. Therefore, it is still challenging to address the aforementioned issues when exploiting social networking data to enhance infection analysis.

In this paper, we propose a Privacy-preserving Infection Analysis approach (PIA) considering social network data associated with health data to infer human-to-human infection spread. This approach employs a privacy-preserving data query method based on conditional oblivious transfer to enable data sharing among different entities and a privacy-preserving classification-based infection analysis method to enable the cloud servers to infer infection spread and preserve health data privacy. The main contributions of this paper are four-fold.

• Firstly, we analyze the spread process of infectious disease with the consideration of user's social contact and health condition. We exploit several key factors of infection, including immunity strength of the susceptible user, infectivity of the infected patient, their contact duration, and the type of contact. We also utilize naive Bayesian classification method to enhance infection analysis with the collaboration of social and health cloud servers.

• Secondly, we propose a privacy-preserving data query method (PPDQ) based on conditional oblivious transfer to allow the authorized entity (i.e., hospital) to access the infected patient's social network data from the social cloud server, but not allow the social cloud server to access and infer any data including patient's identity. Furthermore, this method enables users to grant authorization to hospital, which cannot query any data without user's authorization.

• Thirdly, we propose a privacy-preserving classification-based infection analysis method (PCIA) to prevent user's private social and health data from disclosing to the untrusted health cloud server. The PCIA enables users to encrypt raw data based on homomorphic encryption and send ciphertexts to the cloud server. Then, the health cloud server can infer infection spread during human-to-human contact without learning any user's private information.

• Finally, privacy analysis shows that the proposed approach preserves the privacy of user's health data and social network data, and achieves patient's identity privacy during the query. Furthermore, we conduct the extensive simulation to demonstrate that the PIA exploits the social network data and adjusts to effectively analyze infection spread with acceptable computational overhead.

The remainder of the paper is organized as follows. We review the related works in Section 2. Then, we present the system model and design goals in Section 3. We propose the PIA with details in Section 4 and Section 5. The privacy properties are analyzed in Section 6, and the performance is evaluated in Section 7, respectively. Finally, we conclude the paper in Section 8.

2 Related Works

Social network data analysis has attracted a lot of attentions from both academic and industrial fields as the big volume of social network data are collected for analysis [15], [16]. Some sophisticated machine learning schemes, such as support vector machine, naive Bayesian classification and decision tree based classification [17], are widely applied in practical applications [18]. These schemes usually require the labeled training data set to establish the learning/classification model, which is used to classify the new data. In addition, abnormal event detection is of great importance especially in social network analysis, and requires prior expert knowledge with well-defined models [19], [20]. Due to rarity, unexpectedness and relevance features of abnormal events, Zhang et al. [21] develop a semi-supervised adapted Hidden Markov Model with Bayesian adaptation to adjust abnormal events. It first labels an abnormal event model in an unsupervised pattern from a large volume of ground truth data. An iterative structure is utilized to adapt any emerging abnormal event at each iteration. This framework can address the difficulty in labeling abnormal events and the scarcity of training data [22].

To leverage privacy preservation and data usability [23] for social network data analysis, extensive research efforts have been put in recent years. A variant of "doubly homomorphic" encryption scheme [24] for secure multi-party computation is introduced to perform flexible operations over the encrypted data. With the advanced and efficient homomorphic encryption techniques [25], Graepel et al. [26] propose a machine learning scheme with privacy preservation to outsource the heavy computation tasks to the powerful cloud servers. At the same time, data confidential and user privacy are achieved with the advantages of the adopted leveled homomorphic encryption scheme. This privacy-preserving machine learning scheme mainly solves the privacy issues during the data training phase. To perform both training and learning over encrypted data, Bost et al. [18] develop a set of secure machine learning classification schemes based on leveled fully homomorphic

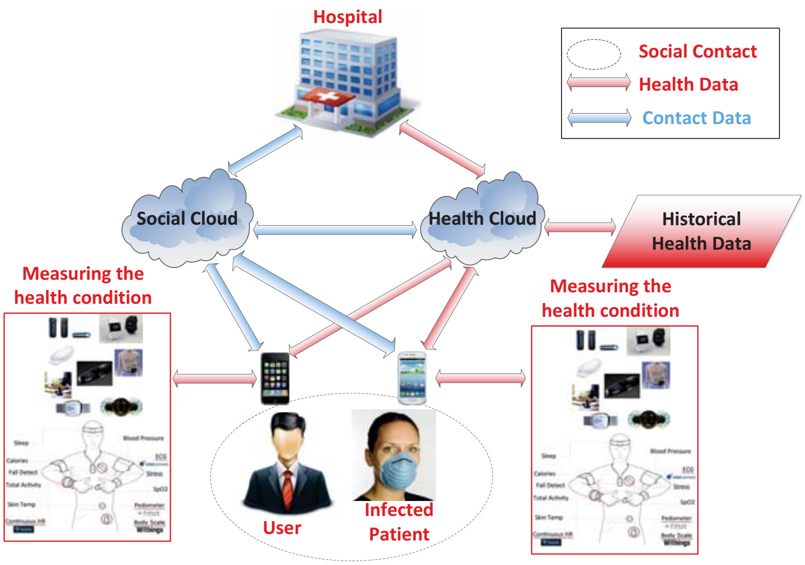


Fig. 1. Infection Analysis System Considering Social Network and Health Data

encryption. In [18], a client performs learning operations with an untrusted server over ciphertexts. In [27], Barni et al. develop a neural network based classification scheme with privacy preservation with linear branching programs to address privacy issues in ECG classification. Samanthula et al. [28] propose a k-nearest neighbor classification algorithm based on Paillier cryptosystem [29] which enables operations over ciphertexts for healthcare systems. In [17], a privacy-preserving clinical decision support system is proposed based on naive Bayesian classification. It first aggregates ground truth data for training, and then enables untrusted cloud servers to perform secure classification algorithm over encrypted data. Users are also allowed to retrieve top-k diagnosis results with their interests and requests. Yuan et al. [30] propose a privacy-preserving back-propagation neural network learning algorithm based on "doubly homomorphic" encryption. This algorithm allows users to send encrypted data to the cloud server, which performs most of the computation tasks without compromising the privacy of user's raw data. Another type of lightweight machine learning is decision tree based classification, which is studied in [31] and developed with privacy protection mechanisms. Recently, Zhou et al. [32] propose a secure text mining scheme, where a privacy-preserving data aggregation method is served as the building block to enable data training in cloud assisted e-healthcare system. Considering the data access problem, Zhou et al. [33] propose a user-controlled multi-level cooperative authentication scheme to protect user's attribute information from disclosing during the data exchange.

However, most of existing works focus on a single cloud platform involving in e-healthcare systems. Due to the unique characteristics of infectious disease, it is necessary to integrate various sources of user's information, such as health and social data for infection analysis. Meanwhile, the large volume of long-lasting health and social data from users pose a big challenge for data management and collaboration in the traditional e-healthcare framework. Therefore, multiple independent cloud servers with different functionalities are involved in our approach to enhance the infection analysis with sufficient knowledge of patients and susceptible users from both health and social perspectives. In addition, data privacy, usability (i.e., secure operations over encrypted data) and efficiency should be considered when designing a novel infection analysis system.

3 System Model and Design Goals

In this section, we propose the infection analysis system model and identify the design goals, respectively.

3.1 System Model

The proposed infection analysis system consists of five entities: trusted authority (TA), users (i.e., data owners), hospital, social cloud server (SC) and health cloud server (HC) as shown in Fig. 1. The system is divided into health domain and social domain according to different types of collected data. Users, HC and hospital have operations on health data in the health domain, while users, SC and hospital (as a query requestor) are involved in the social domain. The details of each entity in the PIA are presented as follows.

- **Trusted Authority (TA)** bootstraps the system, processes user's registration, and generates the certificates for legal user's key generation. Afterwards, TA is not involved in network and users' interactions.
- **Users** first register to the TA and generate valid keys in the initialization phase. They measure their health parameters via wearable devices and periodically send health data to the health cloud server. When user U_i and U_j have contacts with each other, their smartphones record the contact information, such as identity, duration and social relationships, which are sent to the social cloud server.
- **Health Cloud (HC)** has powerful computational and storage capabilities to perform the complicated and time-consuming operations on health data. HC receives health data from users, and training data from medical institutions for analysis.

- **Social Cloud (SC)** is the cloud server dedicated for social network data storage and processing [34], which is similar to HC. SC only operates in social domain.
- **Hospital (*H*)** is the entity to analyze user's infection status. If H diagnoses a user U_i as infected patient, H queries U_i's social network data from SC. Having U_i's social network data, H performs infection analysis with HC to determine whether U_i's encountered users are susceptible or not, and informs users the analysis results.

3.2 Security Model

HC and SC are honest-but-curious entities in the system, i.e., they honestly follow the protocols but are curious about users and other entities' private information. H is trusted by users and has the authorization from users to access their health data stored on HC. However, H is semi-trusted in social domain. If H diagnoses a user U_i as infected patient, U_i grants authorization to H and allows H to access U_i's social data from SC. Otherwise, H is an honest-but-curious entity in social domain and not allowed to access any user's social data in SC except without the authorization.

3.3 Privacy Requirements and Design Goals

Under the honest-but-curious model, user's personal information included social data and health data should be kept confidential towards untrusted and unauthorized entities. The privacy requirements are identified as follows.

(1) ***Health Data Privacy.*** User's health data should be prevented from disclosing to other unauthorized entities, such as HC, SC and any unauthorized user. Particularly, the infected patient's infectivity and other health data are highly privacy sensitive and should not be disclosed to the cloud servers and other users. Furthermore, historical data (i.e., training data set) from medical institutions should also be encrypted in the ciphertext during the classification by HC.

(2) ***Social Data Privacy.*** User's social contact data are part of user privacy. The encountered user's information, such as identities, contact type and duration, should be invisible to SC and other users when the data are stored in SC. Without user's authorization, H should not be able to access this user's social network data as well.

(3) ***Privacy of Susceptible User and Infected Patient.*** Some users may be susceptible to be infected. Before diagnosis, their information, such as identities and health status, should also be protected against SC's inferring. In addition, susceptible user's risk analysis result) should be invisible to HC, SC and any user except authorized hospital. This risk analysis result reflects user's infection status, which is highly sensitive to him.

The proposed system should achieve privacy requirements and computational efficiency simultaneously. On one hand, the proposed system should be able to protect user's social data and health data from disclosing to (or inferring by) untrusted entities. On the other hand, it should take a reasonable computational and communication overheads, which would prolong the system's lifetime and improve user's experiences.

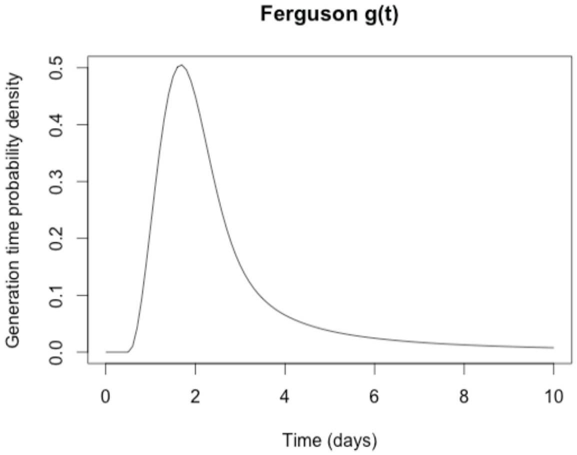


Fig. 2. Infectious Disease Spread Trend [36]

4 Infection Analysis

Infection spread depends on various factors related to health condition and social contact. In this section, we discuss the analytic model on spread of infectious disease and utilize naive Bayesian classification to infer the infection spread.

4.1 Analysis of Infectious Disease Spread

We first propose the analytic model of infectious disease spread. Many infectious diseases, such as acute respiratory diseases, H1N1, measles and flu, can be spread human-to-human via infected droplets during sneezing or coughing, as well as contaminated surfaces and hands. For instance, in a conference environment, Alice has flu and attends the conference where crowd of people are in the same area/room. Alice has many contact with other people such that the flu is likely spread to the contacted people if they do not have sufficient antibody against this type of flu.

(1) ***Factors on Infectious Disease Spread***

The infectious disease spread process between a patient u_a^* and a normal user u_b may be impacted by several factors, i.e., u_a^*'s status (e.g., spread strength), contact duration between u_a^* and u_b, u_b's health condition (e.g., immunity strength).

- ***Infectivity from infected user***: We characterize patient u_a^*'s status in terms of his infectivity $\mathsf{IF}_{u_a^*}$ as a function of the time starting from when u_a^* is a case [35]. The infectivity IF is impacted by the time (t_0) of symptom start when u_a^* is a case, the time (t_1) of infection when u_a^* is a case, and a vector $\mathbf{x}_a^*$ of u_a^*'s personal health status measured by wearable devices (temperature, blood oxygen saturation, etc.) and from hospital including white blood cell and red blood cell content, hemoglobin, etc. Furthermore, symptoms indicate the infectivity to some extent. For example, users have acute respiratory diseases may have at least two symptoms among fever, cough, sore throat, and runny nose. The infectivity is proportional to the strength of these symptoms, but is not proportional to time. As shown in Fig. 2, the generation time of infectious disease is relatively short (only 2 days) [36]. The infectivity (shown as probability density) keeps increasing at the beginning and decreases after the infectious period.

• ***Contact***: According to a recent study [4], sitting next to a patient or being his playmate in a short contact period (e.g., contact lasting minutes with the patient) is not expected to considerably increase the risk of infection. However, a long period contact with patients, such as the structuring of school into classes and grades, would strongly affect spread with an increasing infection risk. The contact duration between the patient u_a^* and normal user u_b can be denoted as $\mathsf{D}_{a,b}$.

Another important feature of user's contact is the contact type $\mathsf{TC}_{a,b}$. We define $\mathsf{TC}_{a,b}$ as 1 = "household" (users living in the same house), 2 = "office" (users in the same office (or classroom in school)), 3 = "department" (uses in the same department (or grade in school)), 4 = "company" (users in the same company (or school)), 5 = "community" (users from the same community, club or social group). The contact type also reflects the social relationship and social-tie strength between contacted users. From "household" to "community", $\mathsf{TC}_{a,b}$ decreases.

The contact information can be bi-directionally captured by wearable devices and smartphones in various ways. User's smartphones can start a Bluetooth discovery program to find the nearby users within a certain range, e.g., 5m or 2m. The contact duration is easy to record by smartphones. Alternatively, GPS and WiFi techniques are possible to measure the location or distance between the contacted users. But this approach is inevitable to face the problem of localization accuracy, especially in the indoor environment or when the accuracy requirement is within meters. Contact type can be captured through the contacted user's social network profiles, such as Facebook, Twitter and WeChat. As the contact information is accumulated at the user side, SC is adopted to help users to store their contact information.

• ***Normal user's health condition***: When a normal user u_b has contact with the patient u_a^*, u_b's health conditions, especially immunity strength, and some other parameters including sleep of quality and physical strength, have impact on disease spread. Let IS_b be the immunity strength of u_b against infectious disease. For simplicity, we consider only one type of infectious disease in the following of this paper.

(2) ***Spread Model of infectious disease***

The infectivity to user u_b during a time period T is

$$\mathsf{S}_b = \sum_{u_a^* \in \mathbb{U}, a \neq b} S_{a^*,b}\left(\mathsf{D}_{a,b}, \frac{1}{\mathsf{TC}_{a,b}}, \mathsf{IF}_{u_a^*}, \frac{1}{\mathsf{IS}_{u_b}}\right). \quad (1)$$

Here, $S_{a^*,b}$ a linear function which denotes the instantaneous infectivity from u_a^* to u_b. $S_{a^*,b}$ is proportional to $\mathsf{D}_{a,b}$, $\frac{1}{\mathsf{TC}_{a,b}}$, $\mathsf{IF}_{u_a^*}$ and $\frac{1}{\mathsf{IS}_{u_b}}$. The infectivity increases with the longer contact duration, the closer social relationship, the higher infectivity from the patient and the lower immunity strength. When a user has contact with multiple patients, the infectivity to this user depends on the contacted patient with the maximum infectivity.

When an infectious disease breaks out in a certain human population, responses in behavior changing according to the outbreak can slow down the progression of the infectious disease to some extent [2]. If a person is aware of the disease in a certain local area or proximity, he would take preventions to considerably reduce his susceptibility. It is important to provide analysis results on the susceptible user's infection.

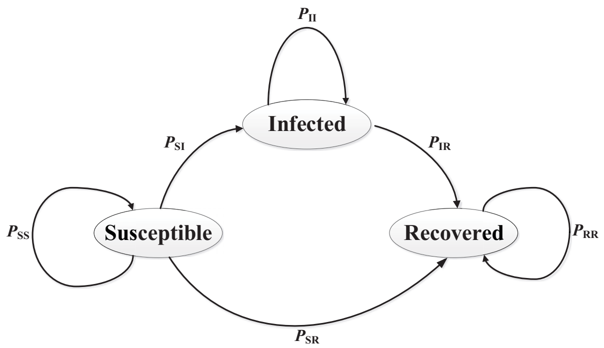


Fig. 3. Infection State of Infectious Disease

Generally, a type of infectious disease has three states on a user U_i, i.e., $\mathbb{S}_i = \{s_{i,1}, s_{i,2}, s_{i,3}\}$. $s_{i,j} \in$ {"Susceptible", "Infected", "Recovered"}, as shown in Fig. 3. The infection process can be formulated as a state transition model, where the "Susceptible" state is the initial state. When a patient recovers from the infectious disease, his immune system can generate antibodies against the pervious infected disease. Similarly to [4], in this paper, we define "Recover" status as the end state.

(3) ***Infection Analysis***

To analyze whether user u_j has a risk to get infected, u_j's health and social data can be considered together to classify if u_j is infected. We utilize naive Bayes classification [37] to analyze infection status. Suppose u_j's data $\mathbf{x} = \{x_1, \cdots, x_l\}$ is an l-dimensional vector, where $x_i \in \mathbb{R}$ and $i \in \{1, \cdots, l\}$. The format of $\mathbf{x}$ is shown in Fig. 4. A classification algorithm $\mathsf{C}(\mathbf{x}, w) : \mathbb{R}^d \mapsto \{\mathbf{c}_1, \cdots, \mathbf{c}_k\}$ takes input as $\mathbf{x}$ and outputs $k^* = \mathsf{C}(w, \mathbf{x}) \in \{1, \cdots, k\}$. Here, k^* is the class (i.e., infection status) to which x corresponds given model w trained by ground truth data. With the abundant health data from hospital, it is feasible to obtain such a model in the PIA. In the model w, each class c_i corresponds to a probability $\{\mathsf{Prob}(C = c_i)\}_{i=1}^{k}$. The j-th element x_j of $\mathbf{x}$ is a and falls into a class c_i with a probability $\mathsf{Prob}(X_j = a | C = c_i)$. Here, A_j is X_j's domain and $a \in A_j$ ($j \in [1, d]$ and $i \in [1, k]$).

The naive Bayes classifier adopts a maximum a posteriori decision rule to select the class with the highest posterior probability as Eqn. 2.

$$\begin{aligned} k^* &= \arg\max_{i \in [k]} \mathsf{Prob}(C = c_i | X = x) \\ &= \arg\max_{i \in [k]} \mathsf{Prob}(C = c_i, X = x) \\ &= \arg\max_{i \in [k]} \mathsf{Prob}(C = c_i, X_1 = x_1, \cdots, X_d = x_d) \end{aligned} \quad (2)$$

$\mathsf{Prob}(X = x)$ is the normalizing factor and deleted given the fixed x according to Bayes theorem.

The Naive Bayes model assumes that $\mathsf{Prob}(C = c_i, X = x)$

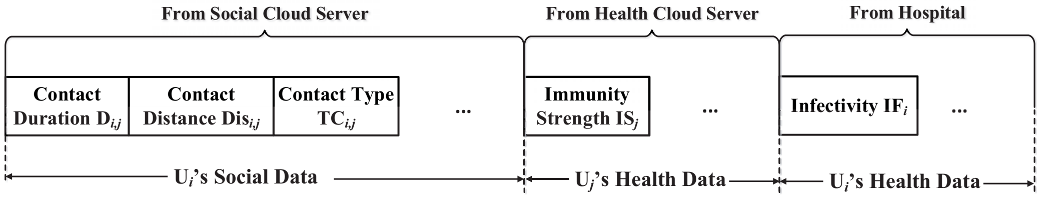


Fig. 4. Input of Bayesian Classification

can be factorized as

$$\begin{aligned}&\mathsf{Prob}(C=c_i,X_1=x_1,\cdots,X_d=x_d)\\&=\mathsf{Prob}(C=c_i)\prod_{j=1}^{d}\mathsf{Prob}(X_j=x_j|C=c_i).\end{aligned}\tag{3}$$

From Eqn. 3, each feature is conditionally independent given the class. The feature value's domain is finite and discrete. The optimal k^* can be selected according to Eqn. 4.

$$\begin{aligned}k^*&=\arg\max_{i\in[k]}\{\log\mathsf{Prob}(C=c_i|X=x)\}\\&=\arg\max_{i\in[k]}\{\log\mathsf{Prob}(C=c_i)\\&\quad+\sum_{j=1}^{d}\log\mathsf{Prob}(X_j=x_j|C=c_i)\}\end{aligned}\tag{4}$$

The class c_{k^*} corresponds to the infection status of user u_j. The integration of social network data and health data includes key factors of infection spread and enhances the traditional infection analysis.

5 Privacy-preserving Infection Analysis Approach

In this section, we exploit social network data to analyze infection spread in a privacy-preserving way. We first provide an overview of the PIA approach. Then, we propose a privacy-preserving data query method and a privacy-preserving classification-based infection analysis method to achieve the design goals.

5.1 Overview of PIA

The PIA adopts naive Bayesian classifier to detect the infected and susceptible users. It consists of four components as shown in Fig. 5. In each component, user's privacy is protected from disclosing to untrusted entities.

(1) ***Health Data Collection***

Users first adopt on-body sensors and wearable devices to measure their health parameters (temperature, heart rate, sleep quality, ECG, exercise statistics and so forth). Before sending the data to HC, users encrypt the measured health data into ciphertexts since these health data are highly privacy-sensitive to users. Finally, the health data are stored in HC.

(2) ***Social Data Collection***

When users contact each other, their smartphones can record the detailed contact information, including identity, duration, social-tie and type of contact, which can be timely uploaded to SC for storage. The smartphones utilize the short range communication techniques (e.g., bluetooth and NFC) and some other built-in sensors (e.g., acoustic sensors) to detect the contact between the smartphone owner and the other user when they are in the physical proximity [38], [34]. The included information is highly private-sensitive and should be invisible to unauthorized entities (e.g., SC).

(3) ***Privacy-preserving Data Query***

The hospital H diagnoses infected disease of patients and determines the infectivity IF as shown in Fig. 2 according to [36]. Then, H informs the infected patients with the diagnosis results. After the diagnosis, H performs a social contact query to SC with the authorization from users. H sends the query request, including infected user u_j's identity associated with the queried contact duration and some other social information, to SC in the ciphertext. SC performs operations on the query request without knowing the query result and feedbacks it to the hospital.

(4) ***Privacy-preserving Infection Analysis***

HC and H compute the contacted user U_i's infection status based on U_i's immunity (measured by U_i), u_j's infectivity, contact duration, contact type and social-tie in a privacy-preserving way. Finally, H sends the analysis results to U_i as the guidelines to treat the potential disease.

5.2 Health Data Collection

To preserve user's health data privacy, users should encrypt their data before sending to the cloud servers. We revisit an RLWE (Ring Learning With Error) based somewhat homomorphic encryption scheme [25] as the preliminary to construct our building block. In the initialization phase, the TA picks the system parameters as follows: 1) a ring $R=\mathbb{Z}[x]/\langle f_\omega(X)\rangle$ where f_ω is ω-th cyclotomic polynomial; 2) an odd positive integer modulus q and a prime $p\ll q$ as the plaintext base; 3) the dimension n and $N=polylog(q,\omega)$; 4) a ring over modulus q is $R_q=R/qR$; and 5) an error distribution χ with small coefficient.

The TA runs a key generation algorithm KeyGen to generate user u's secret key sk_u and public key pk_u. $sk_u=(1,\mathbf{s})\in R_q^{n+1}$, where $\mathbf{s}$ is randomly selected from χ^n. The TA randomly selects $\mathbf{e}=(e_1,\cdots,e_N)\in R^N$ and $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_N)\in R_q^N$. Then, the TA computes $\beta_i=\alpha_i\mathbf{s}+p\cdot e_i \bmod (f_\omega(X),q)$. $pk_u=(\beta_i,-\boldsymbol{\alpha})$.

An encryption algorithm Enc takes input as pk_u and message $M\in R_p$. It makes $\mathbf{m}=(M,0)\in R_q^{N+1}$ and randomly

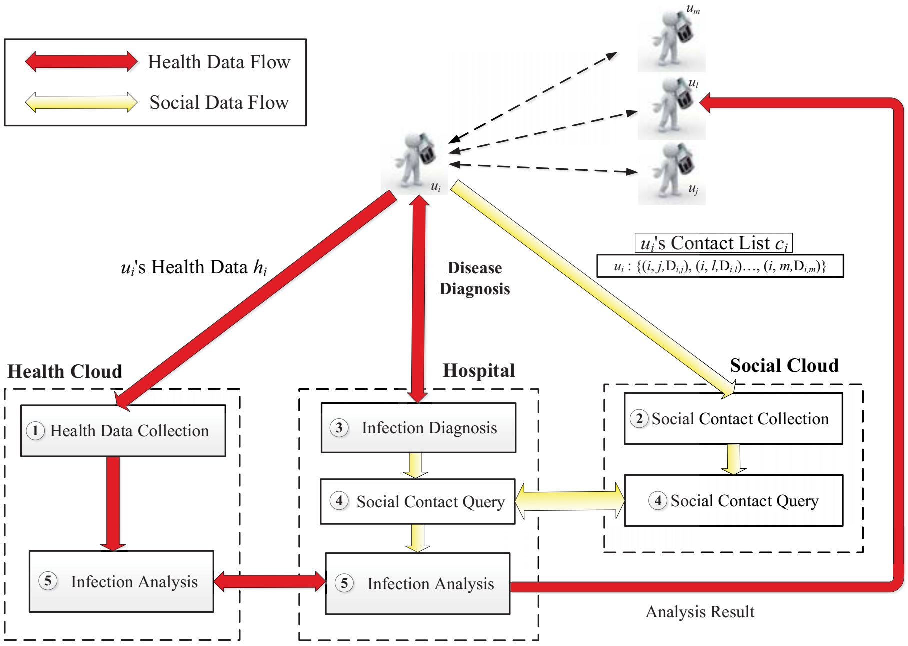


Fig. 5. Illustration of Privacy-preserving Infection Analysis Approach

selects $\mathbf{r} = (r_1, \cdots, r_N) \in R_p^N$. The ciphertext is $\mathsf{CT} = \mathsf{Enc}_{pk_u}(M) = \mathbf{m} + \sum_{i=1}^{N} r_i \cdot pk_u \bmod (f_\omega(X), q) \in R_q \times R_q$.

A decryption algorithm Dec takes input as CT and secret key sk_u, and outputs the message $M = \mathsf{Dec}_{sk_u}(\mathsf{CT})$ as $\langle \mathbf{CT}, sk_u \rangle \bmod (f_\omega(X), p)$. Here, $\langle \mathbf{CT}, sk_u \rangle = \sum_{j=1}^{n+1} \mathbf{CT}(j) \cdot sk_u(j)$ denotes the inner product. We have

$$\begin{aligned}\langle \mathbf{CT}, sk_u \rangle =& M + \sum_{i=1}^{N} r_i \langle sk_u, pk_u \rangle \\ =& M + p \sum_{i=1}^{N} r_i e_i \\ =& M + p \sum_{i=1}^{N} r_i e_i (\text{mod } f_\omega(X), q).\end{aligned} \tag{5}$$

Since $\mathbf{e}$ and $\mathbf{r}$ (i.e., e_i and r_i) are small, $\delta = \sum_{i=1}^{N} r_i e_i (\text{mod } f_\omega(X), q)$ is small such that M can be finally decrypted [25].

This homomorphic encryption scheme can support addition and multiplication operations over ciphertexts. Specifically, the addition of M_1 and M_2 is achieved via component-wise addition of the ciphertexts $\mathsf{Enc}_i(M_1)$ and $\mathsf{Enc}_i(M_2)$. Let $\langle \mathbf{CT}_1, sk_u \rangle = M_1 + p \cdot \delta_1$ and $\langle \mathbf{CT}_2, sk_u \rangle = M_2 + p \cdot \delta_2$. $\langle \mathbf{CT}_1 + \mathbf{CT}_2, sk_u \rangle = (M_1 + M_2) + p \cdot (\delta_1 + \delta_2)$. $M_1 + M_2 = \mathsf{Dec}_i(\mathsf{Enc}_i(M_1) + \mathsf{Enc}_i(M_2))$ if $\delta_1 + \delta_2$ is still small.

To obtain the multiplication of M_1 and M_2, the multiplied ciphertext is $\mathsf{Enc}_i(M_1) \times \mathsf{Enc}_i(M_2)$ as shown in Eqn. 6.

$$\begin{aligned}&\langle \mathbf{CT}_1, sk_u \rangle \times \langle \mathbf{CT}_2, sk_u \rangle \\ &= (M_1 + p \cdot \delta_1) \cdot (M_2 + p \cdot \delta_2) \\ &= M_1 \cdot M_2 + p \cdot (p\delta_1\delta_2 + M_1\delta_2 + M_2\delta_1)(\text{mod } f_\omega(X), q)\end{aligned} \tag{6}$$

If $p\delta_1\delta_2 + M_1\delta_2 + M_2\delta_1$ is small, $M_1 \times M_2 = \mathsf{Dec}_i(\mathsf{Enc}_i(M_1) \times \mathsf{Enc}_i(M_2))$.

With the addition and multiplication over the ciphertext, homomorphic encryption schemes can allow an untrusted entity to perform these operations without knowing secret keys and the content included in the ciphertexts.

When U_i measures his health data h_i, U_i encrypts h_i as $\mathsf{Enc}_i(h_i)$. To enable the hospital (i.e., trusted entity) to access U_i's health data, U_i generates re-encryption key $\widetilde{RK}_{i \to H}$ to transform $\mathsf{Enc}_i(h_i)$ to $\mathsf{Enc}_H(h_i)$ according to [39], [25].

5.3 Social Data Collection

When two users U_i and U_j move in the physical proximity of each other, the contact information, such as contacted users' identities, contact duration, contact type (or social-tie), are recorded by users' smartphones. For example, a Wechat application on smartphones can launch a friend discovery program to find the nearby users (running the same application) and allow them to chat with each other. For simplicity, we assume that user's smartphone measures the social contact of the nearby users as indicated in [38]. We formulate social contact as follows. Let $\mathsf{CI}(i, j)$ denote the contact data between i and j. $\mathsf{CI}(i, j) = (i, j, \mathsf{D}_{i,j}, \mathsf{TC}_{i,j}, \cdots)$. Then, U_i converts $\mathsf{D}_{i,j}$ to a binary vector $\mathbb{D}_{i,j} = \{D_{i,j,1}, D_{i,j,2}, \cdots, D_{i,j,\omega}\}$ where $\omega = \lceil \log l \rceil$ and l is the maximum duration. For example, if users upload their social information to SC every hour, $l = 60$ with minute as the unit of contact time (or $l = 3600$ when using second as unit). The contact duration is a keyword during the query. It is encrypted as $\mathsf{Enc}_i(D_{i,j}) = \{\mathsf{Enc}_i(D_{i,j,1}), \mathsf{Enc}_i(D_{i,j,2}), \cdots, \mathsf{Enc}_i(D_{i,j,\omega})\}$. If U_i grants the authorization of social information query to the hospital, U_i generates re-encryption key $RK^*_{i\to H}$ to transform $\mathsf{Enc}_i(D_{i,j})$ to $\mathsf{Enc}_H(D_{i,j})$.

To make user's uploaded social data invisible to the untrusted SC, these data should be encrypted. Let $\mathbb{G}$ be a cyclic group of order p with generator $g \in \mathbb{Z}_p^*$ [40]. U_i randomly chooses his secret key $SK_i = x_i \in \mathbb{Z}_q$. U_i computes $PK_i = g_i^x$. During the encryption, U_i randomly chooses $r \in \mathbb{Z}_q$, and encrypts $\mathsf{CI}(i,j)$ as $\mathsf{E}_i(\mathsf{CI}(i,j)) = (c_1, c_2) = (g^r \bmod p, \mathsf{CI}(i,j)g^{x_i r} \bmod p)$. To decrypt $\mathsf{E}_i(\mathsf{CI}(i,j))$, the decryptor computes $\mathsf{CI}(i,j) = c_2/(c_1^{x_i})^{-1}$. Finally, U_i sends $\mathsf{E}_i(\mathsf{CI}(i,j))$ and $\mathsf{Enc}_i(D_{i,j})$ to SC.

If U_i grants H the authorization to query U_i's social information in SC, U_i generates the re-encryption key to SC as a proxy to re-encrypt U_i's ciphertext for the hospital. Specifically, U_i splits his secret key x_i into two parts $x_{i,0}$ and $x_{i,1}$ such that $x_i = x_{i,0} + x_{i,1}$ [41], [42]. SC has the re-encryption key $RK_{i\to H} = x_{i,0}$. H receives the decryption key as $x_{i,1}$. To re-encrypt U_i's ciphertext $\mathsf{E}_i(\mathsf{CI}(i,j))$, SC computes $c_2' = c_2/(g^r)^{RK_{i\to H}}$ and outputs the ciphertext as $\mathsf{E}_{i\to H}(\mathsf{CI}(i,j)) = (c_1, c_2')$. Note that H can decrypt U_i's social information by computing $\mathsf{CI}(i,j)g^{x_2 r}/(g^r)^{x_2}$.

5.4 Privacy-preserving Data Query

After making diagnosis of the infected patients, the hospital initiates a query to SC to find the contacted users in a certain period with the infected patients. These users may have potentials to be infected. Since SC is not trusted, the disclosing of users' contact information, e.g., when and where to meet another user, may violate their privacy such that attackers would infer user's habits and preference. In particular, the infected patient's identity is another kind of sensitive information. Imagine that SC knows the hospital queries certain user's social contact data. It is very likely that this queried user either has already been infected or is susceptible. Therefore, it is essential to prevent SC from knowing the query content from the hospital and replied results to the hospital. To protect user's social information from disclosing to SC, the uploaded social contact data should be encrypted. However, it poses a new challenging issue to enable the hospital's oblivious query [43]. To this end, we propose a privacy-preserving data query method (PPDQ) based on conditional oblivious transfer, which allows the hospital to query users' encrypted social contact data in SC without disclosing the query content and results.

The hospital picks the infected patient U_i's identity i and sends $\mathsf{Query}(i,d,s)$ to SC. The hospital receives the query result $\mathsf{Q.Result}(\mathcal{CL}_i)$. Note that $\mathcal{CL}_i = \{\mathsf{CI}(i,j_1), \mathsf{CI}(i,j_2), \cdots, \mathsf{CI}(i,j_m)\}$, where $\mathsf{CI}(i,j_x) = (i, j_x, D_{i,j_x}, ST_{i,j_x})$ ($x \in \{1,\cdots,m\}$) is i's contacted user with $D_{i,j_x} > d$ and $ST_{i,j_x} > s$. For simplicity, we present the details of the query containing identity and contact duration. The other social metrics can be simply extended based on the PPDQ. The hospital requests a range of query user list (including n users) from SC to blind the exact queried user i.

***Step* 1:** The hospital H builds an identity query vector (n-dimension) $I = \{0, 0, \cdots, 0, 1, 0, \cdots, 0\}$, where i-th element of I is 1 and others are 0 (i.e., H queries U_i's data). Then, the hospital converts the minimum contact duration d to a binary vector $\mathbb{D} = \{D_1, D_2, \cdots, D_\omega\}$. Note that $\omega = \lceil \mathrm{log} l \rceil$. The hospital sends $\mathsf{Enc}_H(I)$ and $\mathsf{Enc}_H(d)$ to SC for query. Here, $\mathsf{Enc}_H(I) = \{\mathsf{Enc}_H(I_1), \mathsf{Enc}_H(I_2), \cdots, \mathsf{Enc}_H(I_n)\}$, and $\mathsf{Enc}_H(d) = \{\mathsf{Enc}_H(D_1), \mathsf{Enc}_H(D_2), \cdots, \mathsf{Enc}_H(D_\omega)\}$. In this section, we present the details of how to query 1-of-n users. The PPDQ can be also extended to query k-of-n users.

***Step* 2:** SC holds $e_{i,0} = \mathsf{E}_{i\to H}(\mathsf{CI}(i,j_x))$ and $e_{i,1} = \perp$. Then, SC performs as follows.

a) Compute $\mathsf{Enc}_H(P_y) = \mathsf{Enc}_H(d_y) - \mathsf{Enc}_H(D_{i,j,y})$ for $1 \leqslant y \leqslant \omega$, implying $P_y = d_y - D_{i,j,y}$.

b) Compute $\mathsf{Enc}_H(R_y) = (\mathsf{Enc}_H(d_y) - \mathsf{Enc}_H(D_{i,j,y}))^2$, implying $R_y = (d_y - D_{i,j,y})^2$.

c) Set $\theta_0 = 0$ and compute $\mathsf{Enc}_H(\theta_y) = 2 \cdot \mathsf{Enc}_H(\theta_{y-1}) + \mathsf{Enc}_H(R_y)$, implying $\theta_y = 2 * \theta_{y-1} + R_y$.

d) Choose a random number $r_y \in \mathbb{Z}_p$ and compute $\mathsf{Enc}_H(\beta_y) = \mathsf{Enc}_H(P_y) + \mathsf{Enc}_H(r_y) \times [\mathsf{Enc}_H(\theta_y) - \mathsf{Enc}_H(1)]$, implying $\beta_y = P_y + r_y(\theta_y - 1)$.

e) Choose a random number $\gamma \in \mathbb{Z}_p$ and compute $\mathsf{Enc}_H(\phi_y)$ as

$$\sum_{i=1}^{n} [(e_{i,1} - e_{i,0})\mathsf{Enc}_H(\beta_y) + (e_{i,1} + e_{i,0})\mathsf{Enc}_H(1)] \\ \times \left[\gamma(\mathsf{Enc}_H(I_i)^2 - \mathsf{Enc}_H(I_i)) + \mathsf{Enc}_H(I_i)\right] \\ + \gamma\left(\sum_{i=1}^{n} \mathsf{Enc}_H(I_i) - \mathsf{Enc}_H(1)\right),$$

implying $\phi_y = \sum_{i=1}^{n} (e_{i,1}(\beta_y + 1) + e_{i,0}(1 - \beta_y)) \times (\gamma(I_i^2 - I_i) + I_i) + \gamma \times \left(\sum_{i=1}^{n} I_i - 1\right)$.

Then, SC has a tuple $\mathsf{Enc}_H(\phi) = \langle \mathsf{Enc}_H(\phi_1), \mathsf{Enc}_H(\phi_2), \cdots, \mathsf{Enc}_H(\phi_\omega)\rangle$. SC randomly permutes this tuple and has $\pi(\mathsf{Enc}_H(\phi))$, which is sent to the hospital as the query result.

***Step* 3:** Receiving the tuple from SC, the hospital decrypts the tuple and obtains the effective query result $2e_{i,0}$ if $d < \mathsf{D}_{i,j}$; and $2e_{i,1}$ otherwise. Finally, the hospital decrypts U_i's social information by computing $\mathsf{CI}(i,j)g^{x_2 r}/(g^r)^{x_2}$.

Finally, the hospital can obtain $\mathcal{CL}_i$=$\{(i, j_1, D_{i,j_1}, ST_{i,j_1}), (i, j_2, D_{i,j_2}, ST_{i,j_2}), \cdots, (i, j_m, D_{i,j_m}, ST_{i,j_m})\}$ where $u_{j_1}, u_{j_2}, \cdots, u_{j_m}$ have contact duration ($> d$) with the infected patient U_i.

5.5 Privacy-preserving Classification-based Infection Analysis

We propose a privacy-preserving classification-based infection analysis method (PCIA) to analyze the infection status based on naive Bayesian classification. The input vector includes susceptible user's immune strength, contact information with the patient and this patient's infectivity as indicated in Fig. 4. The infectivity is diagnosed and assigned by the hospital, while the immunity strength is measured by user and stored on HC. H performs PPDQ with HC to retrieve user's health data without directly disclosing any identity and health data to HC. We present details of the key components of the PCIA, including

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2016.2626288, IEEE Transactions on Dependable and Secure Computing

Algorithm 1 Privacy-preserving Comparison Algorithm

```
1: **Input**: $\mathsf{Enc}_H(x)$, $\mathsf{Enc}_H(y)$
2: **Output**: $x > y$
3: HC computes $\mathsf{Enc}_H(a) = \mathsf{Enc}_H(y) + \mathsf{Enc}_H(2^l) - \mathsf{Enc}_H(x)$ and
   randomly selects $r \in (0, 2^{\lambda+l})$. Then, HC computes $\mathsf{Enc}_H(\theta) =
   \mathsf{Enc}_H(a) + \mathsf{Enc}_H(r)$ and sends $\mathsf{Enc}_H(\theta)$ to $H$.
4: $H$ decrypts $\mathsf{Enc}_H(\theta)$ by using $sk_H$, and computes $\eta = \theta \bmod 2^l$.
5: HC computes $\omega = r \bmod 2^l$. Then, HC privately computes $\mathsf{QE}_H(u)$
   with $H$, and obtains $u = 1$ if $\eta < \omega$ according to DGK cryptosystem
   [45].
6: $H$ encrypts $\theta_l$ as $\mathsf{QE}_H(\theta_l)$, which is sent to HC.
7: HC encrypts $r_l$ and computes $\mathsf{QE}_H(\gamma) = \mathsf{QE}_H(u) \cdot \mathsf{QE}_H(\theta_l) \cdot \mathsf{QE}_H(r_l)$.
   Then, HC sends $\mathsf{QE}_H(\gamma)$ to $H$.
8: $H$ decrypts $\gamma$ and finds $\gamma = 0$ if $x > y$; otherwise, $\gamma = 1$.
```

Algorithm 2 Privacy-preserving Argmax Algorithm

```
**Input**: $\mathsf{Enc}(x_1), \cdot, \mathsf{Enc}(x_n)$
**Output**: $\mathsf{Enc}(Max)$
HC adopts a random permutation $\pi$ and computes $\mathsf{Enc}_H(x'_i) =
\mathsf{Enc}_H(x_{\pi(i)})$.
Let $\mathsf{max} = 1$ and $\mathsf{Enc}_H(Max) = \mathsf{Enc}_H(x_{\pi(1)})$
**for** $i = 2 : n$ **do**
    $H$ runs PPC with the result $b_i$ in each iteration. $b_i = 1$ if $Max \leqslant
    a_{\pi(i)}$; otherwise, $b_i = 0$.
    HC selects two random numbers $r_i$ and $s_i \in (0, 2^{\lambda+l})$. Then, HC
    computes $\mathsf{Enc}_H(m'_i) = \mathsf{Enc}_H(Max) + \mathsf{Enc}_H(r_i)$ and $\mathsf{Enc}_H(a'_i) =
    \mathsf{Enc}_H(a_{\pi(i)}) + \mathsf{Enc}_H(s_i)$. Then, $\mathsf{Enc}_H(m'_i)$ and $\mathsf{Enc}_H(a'_i)$ are sent
    to $H$.
    **if** $b_i = 1$ **then**
        $H$ sets $\mathsf{max} = i$, and computes $\mathsf{Enc}_H(v_i) =
        \mathsf{Refresh}(\mathsf{Enc}_H(a'_i))$.
    **else**
        $H$ computes $\mathsf{Enc}_H(v_i) = \mathsf{Refresh}(\mathsf{Enc}_H(m'_i))$
    **end if**
    $H$ sends $\mathsf{Enc}_H(v_i)$ and $\mathsf{Enc}_H(b_i)$ to HC.
    HC computes $\mathsf{Enc}_H(Max) = \mathsf{Enc}_H(v_i) + (\mathsf{Enc}_H(b_i) - \mathsf{Enc}_H(1)) \cdot
    \mathsf{Enc}_H(r_i) - \mathsf{Enc}_H(b_i) \cdot \mathsf{Enc}_H(s_i)$.
**end for**
$H$ sends $\mathsf{Enc}_H(\mathsf{max})$ to HC.
Finally, HC computes the result $\pi^{-1}(\mathsf{max})$.
```

Algorithm 3 Privacy-preserving Classification-based Infection Analysis Algorithm

```
**Input**: $(\mathsf{Enc}(x_1), \cdot, \mathsf{Enc}(x_n))$ from $H$
**Output**: $i^*$
$H$ form a vector $\mathbf{x} = (x_1, x_2, \cdots, x_d) \in \mathbb{Z}^d$ containing $u_b$'s collected
health data related to immunity strength $\mathsf{IS}_b$ and $u_a^*$'s infectivity $\mathsf{IF}_a$
(measured by hospital), contact duration and contact type with the patient
$u_a^*$, i.e., $\mathsf{D}_{a,b}$ and $\mathsf{ST}_{a,b}$ which are queried from SC.
HC sends $\mathsf{Enc}_T(P^*(i))$ and $\mathsf{Enc}_T(P_i^j(x))$ (for all possible $x$ in each
feature), which are sent to $H$.
$H$ re-encrypts $\mathsf{Enc}_T(P^*(i))$ and $\mathsf{Enc}_T(P_i^j(x))$ to $\mathsf{Enc}_{HC}(P^*(i))$ and
$\mathsf{Enc}_{HC}(P_i^j(x))$.
**for** $i = 1 : k$ **do**
    $H$ computes $\mathsf{Enc}_{HC}(\mathsf{Prob}_i) = \mathsf{Enc}_{HC}(P^*(i)) +
    \sum_{j=1}^{d} \mathsf{Enc}_{HC}(P_i^j(x_j))$.
**end for**
$H$ runs the PPAM with HC. $H$ obtains $i^* = \arg\max \mathsf{Prob}_i$.
```

privacy preservation techniques on comparison, argmax and classification.

i) ***Privacy-preserving Comparison (PPC)***

During the comparison, HC compares two ciphertexts of integers x and y encrypted by the hospital H's public key. Let l be the bit length of x and y. Since some operations are on single bit, we adopt Quadratic Residuosity (QR) cryptosystem [44] as the additive homomorphic building block to further improve the computational efficiency. Let QR's plaintext space be $\mathbb{F}_2$ (bits) and $\mathsf{QE}(x)$ is the ciphertext of input bit x. SK_{HC} and PK_{HC} are HC's secret and public keys in QR cryptosystem.

The details can be found in Algorithm 1. HC first injects random number r in the computation of $\mathsf{Enc}_H(x)$ and $\mathsf{Enc}_H(y)$ to blind the comparison results against H. Intuitively, the PPC algorithm checks the most significant bit of $\theta = y + 2^l - x$, indicating whether $x \leqslant y$. In line 5 of Algorithm 1, HC and H privately compute $u = 1$ if $\eta < \omega$ based on DGK cryptosystem [45], which is a practical integer comparison protocol with small plaintext size and ciphertext size. It only requires 5 extra multiplication operations, which improves the algorithm efficiency.

ii) ***Privacy-preserving argmax (PPAM)***

The privacy-preserving argmax algorithm (PPAM) allows HC to output the index of the largest value of $x_1, \cdots, x_n$ encrypted under H's secret key. The PPAM can achieve: 1) H can only learn the index of the largest value but learn nothing else; and 2) H cannot learn the order relations between x_i and x_j. The detailed steps of PPAM are illustrated in Algorithm 3. First, HC adopts a random permutation π to prevent H from learning the order of $\{x_1, \cdots, x_n\}$. With π, HC has $\mathsf{Enc}_H(x'_i) = \mathsf{Enc}_H(x'_{\pi(i)})$. H runs PPC with the result b_i in each iteration (totally n iterations), where $b_i = 1$ if $Max \leqslant a_{\pi(i)}$; otherwise, $b_i = 0$. In each iteration, H can randomize the encryption after determining the maximum of the compared two values. A "refresh" algorithm is introduced to randomize ciphertexts of homomorphic encryption [18]. If the "refresher" knows the secret key, it decrypts the ciphertext and re-encrypts it; otherwise, it multiplies a ciphertext of 0. This "refresh" algorithm is implemented by using re-encryption of homomorphic encryption.

iii) ***Privacy-preserving Classification-based Infection Analysis (PCIA)***

In the privacy-preserving classification-based infection analysis (PCIA) method, H and HC computes user u_b's infectious status according to a training set (model) which can be obtained from the ground truth data (in medical center, institution or government). The training process follows [36]. This training set is encrypted by medical health center (T) and stored in HC for classification. T grants the hospital H the authorization of computation between HC and H. This authorization is enabled by re-encryption of homomorphic encryption. The re-encryption key $\widetilde{RK}_{T \rightarrow HC}$ is assigned to H and allows H to transfer T's ciphertext to HC's domain. Since the input of homomorphic encryption is integer, the log of probability should be converted to integer by multiplying a constant Δ. For simplicity, let $P^*(i) = \lceil \Delta \log \mathsf{Prob}(C = c_i) \rceil$ and $P_i^j(x) = \lceil \Delta \log \mathsf{Prob}(X_j = x | C = c_i) \rceil$ where $x \in D_j$ the domain of x_j. The detailed steps are as follows.

In summary, the PIA provides a privacy-preserving computing framework not only for hospital to analyze the infection status within the contacted population but also prevent (infected and susceptible) user's sensitive information from disclosing.

6 Privacy Discussions

In this section, we discuss the privacy features of the PIA according to the design goals in Section 3.

6.1 Health Data Privacy

We discuss the health data privacy of the PIA in the storage and processing phases. When the health data is stored in HC, a user U_i's health data h_i is invisible to HC due to sematic security of homomorphic encryption [25]. In other words, any adversary holding only the public key and ciphertext of h_i cannot learn any information about h_i. Before sending to HC, h_i is encrypted with U_i's public key. Under the honest-but-curious model, HC cannot decrypt or infer h_i without having U_i's secret key if the ring learning with error (RLWE) assumption holds. The RLWE assumption is that to distinguish the following two distributions is infeasible. The distributions are: 1) a uniform sample $(a_i, b_i) \in R_q^2$; 2) another sample $(a_i, b_i) \in R_q^2$ where we uniformly select $s \in R_q$, then uniformly sample $a_i \in R_q$ and $e_i \in \chi$ to have $b_i = a_i \cdot s + e_i$.

Before the infection analysis, HC re-encrypts $\mathsf{Enc}_i(h_i)$ with $\widetilde{RK}_{i \to H}$ which is the homomorphic re-encryption key to H's domain. Similarly, HC still cannot obtain H's decryption key to know h_i. Meanwhile, the infected patient's infectivity and identity is also encrypted in the ciphertext with H's encryption key. The infected patient's health information is invisible to HC.

In the naive Bayesian classification, each entity's view during the execution and interaction can be simulated according to his input and output. In other words, each entity cannot learn anything except its inputs and outputs, i.e., each party's views generated by a simulator are computationally indistinguishable to his views from the protocol. We show that the PPC, PPAM and PCIA protocols are secure under the honest-but-curious model.

In the PPC protocol, HC's real view is $\mathsf{view}_{HC} = (\mathsf{Enc}_H(x), \mathsf{Enc}_H(y), l, \mathsf{PK}_H, pk_H, r, \mathsf{QE}_H(u), \mathsf{QE}_H(\theta_l))$. We can also build a simulator for HC where the simulator's view is $\mathsf{Sim}_{HC} = (\mathsf{Enc}_H(x), \mathsf{Enc}_H(y), \mathsf{PK}_H, pk_H, \widetilde{r}, \mathsf{QE}_H(\widetilde{\theta_l}))$. Due to the semantic security of the adopted homomorphic encryption scheme, the ciphertexts are indistinguishable. The random number distributions are the same in the real view case and simulation case such that view_{HC} and Sim_{HC} are computationally indistinguishable. Meanwhile, H's real view is $\mathsf{view}_H = (\mathsf{SK}_H, sk_H, l, \mathsf{Enc}_H(\theta), \mathsf{QE}_{HC}(\gamma))$. The view of H's simulator is $\mathsf{Sim}_H = (\mathsf{SK}_H, sk_H, l, \mathsf{Enc}_H(\widetilde{\theta}), \mathsf{QE}_H(\widetilde{\gamma}))$. As the random number r is selected by HC and $\theta = a + r$, θ and $\widetilde{\theta}$ have the same distribution such that they are indistinguishable. Then, $(\mathsf{QE}_H(\theta), \mathsf{QE}_H(\gamma))$ and $\mathsf{QE}_H(\widetilde{\theta}), \mathsf{QE}_H(\widetilde{\gamma})$ are also computationally indistinguishable. H's real view view_H and simulation view Sim_H are also indistinguishable. Therefore, the PPC protocol is secure under the honest-but-curious model.

In the PPAM protocol, HC's real view is $\mathsf{view}_{HC} = (\{\mathsf{Enc}_H(x_i)\}_{i=\{1,\cdots,n\}}, \pi, \mathsf{PK}_H, pk_H; \{r_i, s_i\}_{i=\{1,\cdots,n\}}; \{\mathsf{Enc}_H(v_i), \mathsf{Enc}_H(b_i)\}_{i=\{1,\cdots,n\}}, \pi(\arg\max_{i\in[n]} x_i))$. The simulator's view of HC is $\mathsf{Sim}_{HC} = (\{\mathsf{Enc}_H(x_i)\}_{i=\{1,\cdots,n\}}, \widetilde{\pi}, \mathsf{PK}_H, pk_H; \{\widetilde{r_i}, \widetilde{s_i}\}_{i=\{1,\cdots,n\}}, \{\mathsf{Enc}_H(\widetilde{v_i}), \mathsf{Enc}_H(\widetilde{b_i})\}_{i=\{1,\cdots,n\}}; \arg\max_{i\in[n]} x_i)$. Since the distributions of r_i, s_i and $\widetilde{r_i}, \widetilde{s_i}$ are the same, they are indistinguishable. Due to the semantic security of the homomorphic encryption scheme and the PPC protocol, $\mathsf{Enc}_H(v_i), \mathsf{Enc}_H(b_i)$ and $\mathsf{Enc}_H(\widetilde{v_i}), \mathsf{Enc}_H(\widetilde{b_i})$ are also indistinguishable. In addition, π and $\widetilde{\pi}$ are selected by HC such that they are indistinguishable. Therefore, view_{HC} and Sim_{HC} are indistinguishable. On the other hand, H's real view is $\mathsf{view}_H = (\mathsf{SK}_H, sk_H; \{b_i\}_{i=\{2,\cdots,n\}}; \{\mathsf{Enc}_H(m_i'), \mathsf{Enc}_H(x_i)\}_{i=\{2,\cdots,n\}})$. The view of H's simulator is $\mathsf{Sim}_H = (\mathsf{SK}_H, sk_H; \{b_i\}_{i=\{2,\cdots,n\}}; \{\mathsf{Enc}_H(\widetilde{m_i'}), \mathsf{Enc}_H(\widetilde{x_i})\}_{i=\{2,\cdots,n\}})$ Since the permutation π is a mapping function without changing the order of $\{x_i\}_{i=\{1,\cdots,n\}}$, b_i does not change as well. As r_i, s_i are randomly selected by HC, $\mathsf{Enc}_H(m_i')$ and $\mathsf{Enc}_H(\widetilde{m_i'})$ are indistinguishable. Finally, H's real view view_H and simulation view Sim_H are indistinguishable. Therefore, the PPAM protocol is secure under the honest-but-curious model.

In the PCIA protocol, HC cannot view anything other than the inputs since the PPC and PPAM protocols are both secure under the honest-but-curious model. H's real view is $\mathsf{view}_H = (\mathsf{SK}_H, sk_H, \{x_i\}_{i=\{1,\cdots,n\}}; \{\mathsf{Enc}_H(P_i^j)\}_{i=\{1,\cdots,n\};j=\{1,\cdots,d\}}, \mathsf{Enc}_H(P^*), i^*)$. The view of H's simulator is $\mathsf{Sim}_H = (\mathsf{SK}_H, sk_H, \{x_i\}_{i=\{1,\cdots,n\}}; \{\mathsf{Enc}_H(\widetilde{P_i^j})\}_{i=\{1,\cdots,n\};j=\{1,\cdots,d\}}, \mathsf{Enc}_H(\widetilde{P^*}), \widetilde{i^*})$. Due to the semantic security of the homomorphic encryption scheme, PPC and PPAM protocols, view_H and Sim_H are indistinguishable. Therefore, the PCIA protocol is secure under the honest-but-curious model.

6.2 Social Data Privacy

User's social contact information $\mathsf{CI}(i, j)$ is encrypted by U_i with his public key. Without U_i's secret key, SC cannot decrypt and have the plaintext if the decisional Diffie-Hellman problem is hard in $\mathbb{G}$. Therefore, when the social data are stored on SC, no private information of users can be disclosed to SC.

As the hospital H is an honest-but-curious entity in social domain, it follows the protocol without maliciously querying user's social data in SC. Furthermore, the diagnosis from the hospital provides the second-level decryption key for the hospital to decrypt the plaintext of U_i's social contact information. Users are able to grant social contact information access permission by issuing re-encryption key $\widetilde{RK}_{i \to H}$ to allow SC to re-encrypt $\mathsf{CI}(i, j)$ to the hospital's domain. Without the permission, the hospital still cannot decrypt to have $\mathsf{CI}(i, j)$, even though it can obtain the query results from SC. Note that re-encryption is unidirectional such that the users cannot recover the hospital's secret key to decrypt other user's social information. The infected patient's identity is also protected against HC during infection analysis. Therefore, the patient's identities and contact information, including contacted users and duration, are protected against SC. The hospital can only obtain user's social contact information after he is diagnosed as infected.

6.3 Privacy of Patient and Susceptible User

Susceptible user's identity and analysis results can be invisible to HC, SC and any other unauthorized entities. When a patient is diagnosed, the hospital H sends social data query request to SC. During the social data query process, SC learns nothing except that n users are involved in the hospital's query request. But SC cannot know which one user (or k-of-n users) can be queried if PPDQ method is semantic secure under the honest-but-curious model. We show the semantic security of PPDQ method as follows. H's query request $\mathsf{Query}(i, d, s)$ and query result $\mathsf{Q.Result}(\mathcal{CL}_i)$ are privacy-preserving against SC because the adopted homomorphic encryption scheme and ElGamal cryptosystem are semantic secure under the honest-but-curious model. Without the secret key, $\mathsf{Query}(i, d, s)$ and $\mathsf{Q.Result}(\mathcal{CL}_i)$ are invisible to SC. SC's real view is $\mathsf{view}_{SC} = (\mathsf{Enc}_H(d), \mathsf{Enc}_H(I); r_y(1 \leqslant y \leqslant \omega), \mathsf{Enc}_H(\phi))$. The simulator's view of SC is $\mathsf{Sim}_{SC} = (\mathsf{Enc}_H(d), \mathsf{Enc}_H(I); \widetilde{r_y}(1 \leqslant y \leqslant \omega), \widetilde{\mathsf{Enc}_H(\phi)})$. Since the distributions of r_y and $\widetilde{r_y}$ are the same, they are computational indistinguishable. Due to the semantic security of the homomorphic encryption scheme and the PPC protocol, $\mathsf{Enc}_H(\phi)$ and $\widetilde{\mathsf{Enc}_H(\phi)}$ are also indistinguishable. Therefore, SC's real view view_{SC} and simulation view Sim_{SC} are indistinguishable. The identity query vector I can bound the maximum number of H's queried users. H can decrypt the valid result only if $\sum_{i=1}^{n} \mathsf{Enc}_H(I_i) - \mathsf{Enc}_H(1) = 0$. Therefore, the PPDQ is secure under the honest-but-curious model. PPDQ can be secure performed between HC and H when H retrieves users' health data from HC.

Only infected patients grant social data access to H after they are diagnosed in the hospital. Then, the infected patient's decryption key for his re-encrypted ciphertexts is sent to H such that H can decrypt patient's social data after the query. If H arbitrarily builds identity query vector I, H cannot find any valid information due to the semantic security of ElGamal cryptosystem.

7 Performance Evaluation

In this section, we evaluate the performance of the PIA with respect to simulation and computational overhead.

7.1 Simulations

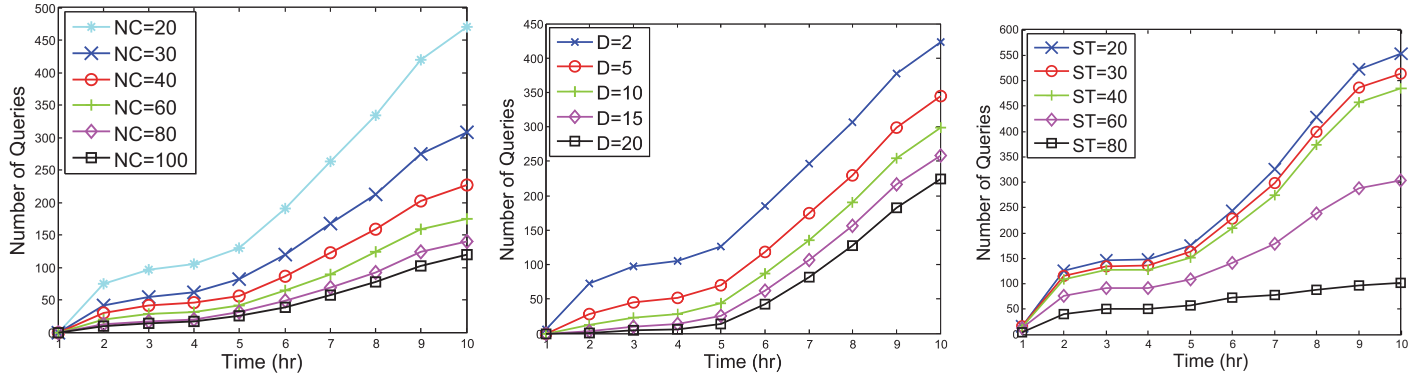
We conduct extensive simulation based on Infocom06 data set [38], which contains 78 mobile users in a conference. Each user takes a portable device with Bluetooth proximity discovery program to find the nearby users. The social-tie (used to reflect contact type) is also obtained according to user's interactions in the data set. We use this scenario to simulate the infectious disease spread under an indoor environment. In this simulation, we randomly select 8 infected patients with a random assigned infectivity value ranging from 50 to 100. We also set user's immunity strength similarly in the range of $[50, 100]$.

In the simulation, we aim to show the trend of the social characteristic impact other than quantifying the formula between immunity strength and infectivity. The hospital or users can define thresholds to trigger queries, where we consider the thresholds of contact number NC, contact duration D and social-tie ST as shown in Fig. 6. From Fig. 6(a), we can see that the number of queries decreases with the increasing threshold contact number. When NC is small, e.g., 20 or 30, more queries are triggered since the PIA provides a conservative strategy to include more queries. As shown in Fig. 6(b), the decreased duration threshold results in the increasing number of queries since the longer contact between the infected patient and normal users could increase the infection risk of the normal users. In Fig. 6(c), the PIA operates with a conservative strategy as ST is small. But the number of queries does not vary too much when ST is from 20 to 40. The reason is that a higher social-tie in a certain range (e.g., in a low level from 20 to 40) may not indicate frequent contacts which are the key factor to accumulate the infection spread [2]. When ST keeps increasing, it shows significant impact on the number of queries. This result also validates the point from [2] that the social relationship is an important factor to influence the spread process of infectious disease and the close relationships (e.g., students in the same class, or families) may cause severe infection spread. By adjusting ST, the PIA can efficiently notify the people with high social-ties to take actions to prevent the infection spread from human-to-human contact. Therefore, the above results validate the trend in Eqn. 1 and show that the PIA is effective in responding to the spread of infectious disease.

7.2 Computational Performance

We use the acute inflammations data set [46] including 120 instances with attributes (i.e., patient's temperature, lumbar pain, urine pushing, micturition pains, urethra status) and corresponding decisions (i.e., inflammation of urinary bladder, and nephritis of renal pelvis origin). We first test the accuracy of the PIA. The total 59 instances with inflammation of urinary bladder and 50 instances with nephritis of renal pelvis origin are all detected. But the PIA detects 47 non-inflammation instances and 59 non-nephritis ones. The accuracy towards individual decision is 88.33% and 90.83%, respectively.

To demonstrate the advantages of using social data for infection analysis, we generate a data set including the contact information from the real world human trace and synthetic health data. In this data set, each instance contains: contact duration, social-tie, immunity strength, infectivity and infection status. According to [36], we use the 1/4 data set (corresponding to the first day of the conference) to label the training set including 100 instances. We generate 200 input instances with the randomly selected health data (i.e., immunity strength, infectivity and infection status according to [36]) for a baseline classification scheme that only has health data to analyze the infection status. Note that we only label "Susceptible" and "Recovered" in the infection status since we focus on the analysis of infection spread. Meanwhile, we generate 200 instances including contact information from the other 3/4 data set of the real world human trace and the same health data used in the baseline classification scheme. In the 200-instance data set, the number of "Susceptible"



(a) Impact of Contact Number

(b) Impact of Contact Duration

(c) Impact of Social-Tie

Fig. 6. Impact of Social Characteristics

TABLE 1
Infection Analysis Comparison

	PIA	Baseline Scheme
“Susceptible”	103/120 (85.83%)	71/120 (59.16%)
“Recovered”	73/80 (91.25%)	62/80 (77.5%)
Overall	176/200 (88%)	133/200 (66.5%)

and “Recovered” is 120 and 80, respectively. As shown in Table 1, the PIA detects 103 “Susceptible” instances and 73 “Recovered” ones, while the baseline scheme detects 71 and 62. Therefore, the integrated social data has the advantages of analyzing infection spread.

With respect to the computational cost of PCIA, we conduct the experiment under a homomorphic encryption library HELib [47] on an Intel Core i5 2.7GHz machine with 4GB RAM to test the computational running time of the proposed methods based on Infocom06 trace. It achieves 80 bits of security with the parameter settings. To mimic the real network environment, we set the communication overhead as 30ms during each interaction of different entities (similar to [18]). In the PPC, H takes 42.94 ms, while HC takes 65.674 ms. In the PPAM, H takes 6.350 s, while HC takes 12.741 s. To perform the PCIA, H takes 7.016 s, while HC takes 24.282 s. Therefore, we can see that HC takes over the majority of the computational overhead since HC has powerful computational capability. The overhead for H is not high, compared with [18] in Table 2.

We also test the running time of the PPDQ with HELib and Crypto++ [48]. We set $l = 1024$, $\omega = 10$ and $n = 78$. H takes 329.234 ms to generate the query to SC and retrieve the results, while SC takes 6.487 s to return the query results. The majority of computational overhead is at the SC side.

8 Conclusions

In this paper, we have proposed a human-to-human infection analysis approach by utilizing social network data and health data to enhance infection analysis without privacy leakage. First, we have analyzed the infectious disease spread process and adopted naive Bayesian classification to detect user's infection status. Furthermore, we have exploited social cloud server to collect users' social networking data, and relied on health cloud server to process/classify users' health data. We have proposed a privacy-preserving data query method to enable hospital to query infected patient's social contacts without allowing the social cloud server to infer the patient's identity and contact details. We have also proposed a privacy-preserving classification-based infection analysis method to perform infection analysis over the encrypted social and health data on the health cloud server. Performance evaluation demonstrates that the PIA can enhance infection analysis efficiency with acceptable overhead. For the future work, we will develop deep learning algorithms for the PIA to perform the comprehensive analysis.

References

[1] “The chief public health officer's report on the state of public health in canada (infectious disease — the never-ending threat),” 2013. [Online]. Available: http://www.phac-aspc.gc.ca/cphorsphc-respcacsp/2013/infections-eng.php

[2] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, “Multiscale Mobility Networks and The Spatial Spreading of Infectious Diseases,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21 484–21 489, 2009.

[3] Z. Sun, F. Wang, and J. Hu, “LINKAGE: An Approach for Comprehensive Risk Prediction for Care Management,” in *Proc. of SIGKDD*, 2015, pp. 1145–1154.

[4] S. Cauchemez, A. Bhattarai, T. Marchbanks, R. Fagan, S. Ostroff, N. Ferguson, D. Swerdlow, S. Sodha, M. Moll, and F. Angulo, “Role of Social Networks in Shaping Disease Transmission During a Community Outbreak of 2009 H1N1 Pandemic Influenza,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 7, pp. 2825–2830, 2011.

[5] K. Zhang, K. Yang, X. Liang, Z. Su, X. Shen, and H. Luo, “Security and privacy for mobile healthcare networks — from quality-of-protection perspective,” *IEEE Wireless Communications*, vol. 22, no. 4, pp. 104–112, 2015.

TABLE 2
Computation Overhead Comparison

	PPC	[18]	PPAM	[18]	PCIA	[18]
H	42.94 ms	53.18 ms	6.350 s	10.370 s	7.016 s	19.311 s
HC	65.674 ms	74.83 ms	12.741 s	18.694 s	24.282	27.595 s

[6] M. Bhuiyan, G. Wang, J. Wu, J. Cao, X. Liu, and T. Wang, “Dependable Structural Health Monitoring Using Wireless Sensor Networks,” *IEEE Transactions on Dependable and Secure Computing*, to appear.

[7] L. Guo, C. Zhang, and Y. Fang, “A Trust-Based Privacy-Preserving Friend Recommendation Scheme for Online Social Networks,” *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 4, pp. 413–427, 2015.

[8] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. Hassan, A. AlElaiwi, and M. Alrubaian, “A performance evaluation of machine learning-based streaming spam tweets detection,” *IEEE Transactions on Computational Social Systems*, vol. 2, no. 3, pp. 65–76, 2015.

[9] B. J. Kolowitz, G. R. Lauro, J. Venturella, V. Georgiev, M. Barone, C. Deible, and R. Shrestha, “Clinical social networkinga new revolution in provider communication and delivery of clinical information across providers of care?” *Journal of Digital Imaging*, vol. 27, no. 2, pp. 192–199, 2014.

[10] K. Zhang, X. Liang, R. Lu, and X. Shen, “Exploiting multimedia services in mobile social network from security and privacy perspectives,” *IEEE Communications Magazine*, vol. 52, no. 3, pp. 58–65, 2014.

[11] R. Yu, J. Kang, X. Huang, S. Xie, Y. Zhang, and S. Gjessing, “MixGroup: Accumulative Pseudonym Exchanging for Location Privacy Enhancement in Vehicular Social Networks,” *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 1, pp. 93–105, 2016.

[12] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, “Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 1, pp. 131–143, 2013.

[13] M. Barreno, B. Nelson, R. Sears, A. Joseph, and J. Tygar, “Can Machine Learning Be Secure?” in *Proc. of ASIACCS*, 2006, pp. 16–25.

[14] C. Gentry and S. Halevi, “Implementing gentry’s fully-homomorphic encryption scheme,” in *Proc. of EUROCRYPT*, 2011, pp. 129–148.

[15] M. Dong, T. Kimata, K. Sugiura, and K. Zettsu, “Quality-of-Experience (QoE) in Emerging Mobile Social Networks,” *IEICE TRANSACTIONS on Information and Systems*, vol. E97-D, no. 10, pp. 2606–2612, 2014.

[16] K. Wei, M. Dong, K. Ota, and K. Xu, “CAMF: Context-Aware Message Forwarding in Mobile Social Networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2178–2187, 2015.

[17] X. Liu, R. Lu, J. Ma, L. Chen, and B. Qin, “Privacy-Preserving Patient-Centric Clinical Decision Support System on Naive Bayesian Classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 2, no. 20, pp. 1261–1273, 2016.

[18] R. Bost, R. Popa, S. Tu, and S. Goldwasser, “Machine Learning Classification over Encrypted Data,” in *Proc. of NDSS*, 2015, pp. 1–14.

[19] C. Stauffer and W. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[20] H. Zhong, J. Shi, and M. Visontai, “Detecting Unusual Activity in Video,” in *Proc. of CVPR*, 2004, pp. 819–826.

[21] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, “Semi-Supervised Adapted HMMs for Unusual Event Detection,” in *Proc. of CVPR*, 2005, pp. 611–618.

[22] K. Zhang, X. Liang, R. Lu, K. Yang, and X. Shen, “Exploiting mobile social behaviors for sybil detection,” in *Proc. of IEEE INFOCOM*, 2015, pp. 271–279.

[23] R. Rivest, “Cryptography and machine learning,” *Lecture Notes in Computer Science*, vol. 739, pp. 427–439, 1993.

[24] Boneh, Goh, and Nissim, “Evaluating 2-DNF Formulas on Ciphertexts,” in *Proc. of TCC*, vol. 2, 2005.

[25] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, “(leveled) fully homomorphic encryption without bootstrapping,” in *Proc. of ITCS*, 2012, pp. 309–325.

[26] T. Graepel, K. Lauter, and M. Naehrig, “ML Confidential: Machine Learning on Encrypted Data,” in *Proc. of ICISC*, vol. 7839, 2012, pp. 1–21.

[27] M. Barni, P. Failla, R. Lazzeretti, A.-R. Sadeghi, and T. Schneider, “Privacy-Preserving ECG Classification With Branching Programs and Neural Networks,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 452–468, June 2011.

[28] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, “k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 5, pp. 1261–1273, May 2015.

[29] Paillier and Pointcheval, “Efficient public-key cryptosystems provably secure against active adversaries,” in *Proc. of ASIACRYPT*, 1999, pp. 1–13.

[30] J. Yuan and S. Yu, “Privacy Preserving Back-Propagation Neural Network Learning Made Practical with Cloud Computing,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 212–221, 2014.

[31] P. Fong and J. Weber-Jahnke, “Privacy Preserving Decision Tree Learning Using Unrealized Data Sets,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 2, pp. 353–364, 2012.

[32] J. Zhou, Z. Cao, X. Dong, and X. Lin, “PPDM: Privacy-preserving Protocol for Dynamic Medical Text Mining and Image Feature Extraction from Secure Data Aggregation in Cloud-assisted e-Healthcare Systems,” *IEEE Journal of Selected Topics in Signal Processing*, to appear.

[33] J. Zhou, X. Lin, X. Dong, and Z. Cao, “PSMPA: Patient Self-Controllable and Multi-Level Privacy-Preserving Cooperative Authentication in Distributedm-Healthcare Cloud Computing System,” *IEEE Transactions on Parallel and Distribited Systems*, vol. 26, no. 6, pp. 1693–1703, 2015.

[34] A. Mohaien, D. Kune, E. Vasserman, M. Kim, and Y. Kim, “Secure Encounter-Based Mobile Social Networks: Requirements, Designs, and Tradeoffs,” *Dependable and Secure Computing, IEEE Transactions on*, vol. 10, no. 6, pp. 380–393, 2013.

[35] C. Fraser, S. Riley, R. Anderson, and N. Ferguson, “Factors that Make an Infectious Disease Outbreak Controllable,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 16, pp. 6146–6151, 2004.

[36] N. Ferguson, D. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. Burke, “Strategies for containing an emerging influenza pandemic in southeast asia,” *Nature*, vol. 437, no. 7056, pp. 209–214, 2005.

[37] A. McCallum and K. Nigam, “A comparison of event models for naive bayes text classification,” in *AAAI-98 workshop on learning for text categorization*, vol. 752, 1998, pp. 41–48.

[38] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, “CRAWDAD trace cambridge/haggle/imote/infocom (v. 2006-01-31),” Jan. 2006.

[39] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in *Proc. of STOC*, 2009, pp. 169–178.

[40] T. ElGamal, “A public key cryptosystem and a signature scheme based on discrete logarithms,” in *Advances in cryptology*, 1985, pp. 10–18.

[41] A.-A. Ivan and Y. Dodis, “Proxy cryptography revisited,” in *Proc. of NDSS*, 2003, pp. 1–20.

[42] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, “Improved proxy re-encryption schemes with applications to secure distributed storage,” *ACM Transactions on Information and System Security*, vol. 9, no. 1, pp. 1–30, 2006.

[43] I. Blake and V. Kolesnikov, “Strong Conditional Oblivious Transfer and Computing on Intervals,” in *Proc. of ASIACRYPT*, 2004, pp. 515–529.

[44] S. Goldwasser and S. Micali, “Probabilistic encryption and how to play mental poker keeping secret all partial information,” in *Proc. of STOC*, 1982, pp. 365–377.

[45] I. Damgard, M. Geisler, and M. Kroigard, “Homomorphic encryption and secure comparison,” *International Journal of Applied Cryptography*, vol. 1, pp. 22–31, 2008.

[46] “Acute Inflammations Data Set,” 2009. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Acute+Inflammations

[47] "HELib," 2013. [Online]. Available: http://shaih.github.io/HElib/index.html

[48] "Crypto++," 2015. [Online]. Available: https://www.cryptopp.com/

Kuan Zhang (IEEE S'13) received the B.Sc. degree in Electrical and Computer Engineering and the M.Sc. degree in Computer Science from Northeastern University, China, in 2009 and 2011, respectively. He is currently working toward a Ph.D. degree at the Broadband Communications Research (BBCR) Group, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include security and privacy for mobile social networks, e-healthcare system, and cloud computing.

Xiaohui Liang (IEEE M'15) received his Ph.D. degree at the department of Electrical and Computer Engineering of the University of Waterloo, and the master and the bachelor degrees at Computer Science department of the Shanghai Jiao Tong University. He was also a postdoctoral researcher at the Department of Computer Science, Dartmouth College, NH, USA. Since 2015, he has been an assistant professor of the Computer Science Department at the University of Massachusetts Boston. His research interests include security, privacy, and trustworthiness in medical cyber physical systems, cyber security for mobile social networks, and applied cryptography.

Jianbing Ni received his master degree and bachelor degrees from University of Electronic Science and Technology of China in 2014 and 2011, respectively. Currently, he is pursuing the Ph.D. degree at the Broadband Communications Research (BBCR) Group, Department of Electrical and Computer Engineering, University of Waterloo. His research interests include the security and privacy for crowdsouring, vehicular ad hoc network, cloud computing and fog computing.

Kan Yang received his B. Eng. degree from University of Science and Technology of China in 2008 and his PhD degree from City University of Hong Kong in August 2013. He is currently a postdoctoral fellow of electrical and computer engineering department at University of Waterloo, Canada. He was a visiting scholar in State University of New York at Buffalo in 2012. His research interests include Cloud Security, Big Data Security, Cloud Data Mining, Cryptography, Social Networks, Wireless Communication and Networks, Distributed Systems etc.

Xuemin (Sherman) Shen (IEEE M'97-SM'02-F'09) received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982, and the M.Sc. and Ph.D. degrees from Rutgers University, Newark, NJ, USA, in 1987 and 1990, respectively, all in electrical engineering.

Dr. Shen is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He was the Associate Chair for Graduate Studies from 2004 to 2008. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks.

Dr. Shen was a recipient of the Excellent Graduate Supervision Award in 2006 and the Outstanding Performance Award in 2004, 2007, 2010, and 2014 from the University of Waterloo; the Premier's Research Excellence Award in 2003 from the province of ON; and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He served as the Technical Program Committee Chair/Co-Chair for ACM MobiHoc'15, IEEE Infocom'14, IEEE VTC'10 Fall, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07. He also serves/served as the Editor-in-Chief for IEEE Network, Peer-to-Peer Networking and Application, and IET Communications. Dr. Shen is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.