Research Article

Privacy-Preserving Health Data Collection for Preschool Children

Shaopeng Guan,1,2 Yuan Zhang,2 and Yue Ji3

1 Key Laboratory of Intelligent Information Processing in Universities of Shandong (Shandong Institute of Business and Technology), Yantai 264005, China
2 State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China
3 Preschool Education Department, Nanjing Normal University, Nanjing 210097, China

Correspondence should be addressed to Yue Ji; yue.ji@njnu.edu

Received 21 August 2013; Accepted 23 September 2013

Copyright © 2013 Shaopeng Guan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of network technology, more and more data are transmitted over the network and privacy issues have become a research focus. In this paper, we study the privacy in health data collection of preschool children and present a new identity-based encryption protocol for privacy protection. The background of the protocol is as follows. A physical examination for preschool children is needed every year out of consideration for the children’s health. After the examination, data are transmitted through the Internet to the education authorities for analysis. In the process of data collection, it is unnecessary for the education authorities to know the identities of the children. Based on this, we designed a privacy-preserving protocol, which delinks the children’s identities from the examination data. Thus, the privacy of the children is preserved during data collection. We present the protocol in detail and prove the correctness of the protocol.

1. Introduction

With computers and networks having become an important tool in everyday life, more and more data need to be transmitted through networks. Meanwhile, privacy issues have drawn public attention. How to protect privacy in a network environment has become a research focus in the field of computer network.

Privacy, broadly speaking, refers to private data held by organizations or individuals, which are confidential to others. For individuals, private information such as personal identification, physical condition, and geographical location is all private [1]. The spread of private information will cause a lot of negative consequences, even leading to crimes. Therefore, privacy-preserving technology becomes an important research direction. At present, researches of privacy-protection technology in the network include at least the following areas.

Privacy Protection in Wireless Sensor Networks. Wireless sensor networks have broad application prospects in the fields of environmental monitoring, health care, national defense, and so on. However, in practical applications, wireless sensor networks are facing a serious risk of data disclosure or tampering that will lead to serious consequences [2–5]. For example, in the field of military, data collected by wireless sensor networks often contain important intelligence information which, if disclosed or tampered with, will pose a serious threat or military missteps. The privacy-protecting technology is an indispensable part of wireless sensor networks [6–10].

Privacy Preserving-Data Mining. Data mining is the most important knowledge discovery tool in today’s society. It can reveal the hidden rules behind large amounts of data for people. However, data sources used for data mining also contain a lot of individual privacy, business intelligence or government secrets. In the data mining process, if the data are used arbitrarily without any restraint, personal privacy and confidential information will be disclosed, and thereby people’s daily lives and even social stability will be seriously affected [11–15]. It is a dilemma to pick up potential and valuable knowledge from the massive amounts of data in data mining and in the meantime preserve privacy. The ideal solution is to transform the raw data, and then prevent the
direct and indirect access to private information, while the mining algorithms are still able to get from the converted data almost the same information and knowledge as those from the raw data [16–20].

Privacy Studies for Medical and Health Information. In the field of medicine, medical treatments and results must be based on the patients’ privacy. With the application of information technology in the medical field, electronic medical records (EMRs) have become the main carrier of medical information. EMRs are prevailing in medical institutions because of their large storage capacity, saving of resources, convenient query, and good sharing of information, which improve the efficiency of diagnosis and treatment [21]. However, since EMRs contain a lot of patients’ privacy and are easy to copy and spread, privacy protection is significant in the field of medical and health information [22–24].

In addition to the above, as new network applications emerge, some new privacy issues also need to be addressed. For example, in fields such as social network, data publishing, cloud computing, and the Internet of Things, privacy has attracted people’s attention [25–28]. In recent years, location privacy in the mobile network also become a highlight as it no longer contains personal information and therefore achieves the protection of individual privacy. Most of the existing privacy-protecting technologies of the raw data are based on anonymous method. Anonymizing the raw data will inevitably result in loss of information. Therefore, the research work is focused on finding the tradeoff between the availability of data and privacy protection [34–37].

Access Control of EMRs. Using a centralized management of rights, the access control technology is a defensive measure against unauthorized use of data. Its basic objective is to control access rights of users to EMRs or medical information system and thereby ensure that the medical data are used under authorization. Access control is an important measure to protect electronic medical data in information systems, which determines who can access the system and how the data are used. Appropriate access controls can prevent unauthorized users from making accidental or inadvertent access to data; however, the implementation of access control is complex, and the adjustment and management of rights are difficult [38–41].

Privacy-Protecting Medical Information System. In addition to the above privacy-protecting technologies, scholars designed some privacy-protecting medical information systems. In [42], Jieun Song and Myungae Chung put forward a safe framework of health privacy for environmental service model. The system includes authentication, access control and privacy-protecting service. In [43], Gardner and Xiong constructed an identity-conversion system to protect the health information of patients. The system uses conditional random fields method to extract identity properties from unstructured data and conduct identity conversion by $k$-anonymous method. In [44], Lin et al. proposed a privacy protecting scheme for electronic health systems and proved by formal reasoning that the scheme is able to protect medical privacy and context information simultaneously. However, till now, there is no perfect system architecture of privacy protection.

On the whole, privacy-protecting technologies in the field of medicine and health have made considerable progress. However, there are still some problems with the existing technologies. For example, security assumptions in the models are too strong to be adopted to the real scenario. In addition, existing privacy-protecting schemes have no universality. Every scheme is only for a specific situation or a specific privacy issue. As far as the privacy in the health data collection for preschool children in this paper is concerned, no existing schemes can be directly adopted. Thus, we design an identity-based encryption protocol for the privacy protection.

2. Related Works

With the application of information technology in the field of medicine and health, the privacy issues also begin to grasp people’s attention. At present, research on EMRs focuses on three areas: privacy protection of raw data, access control of EMRs, and privacy-protecting medical information system [31–33].

Privacy Protection of Raw Data. Privacy protection of raw data refers to the fact that some technologies such as interference or anonymity are adopted to process raw data and form a new data set before the raw data are provided to others. After transformation, the new data set maintains the same distribution characteristics as the raw data, while
a trusted party, called the certification authority (CA), is responsible for authenticating and issuing the corresponding public key certificate of users. The public key certificate binds the identity of a user with its public key. In this kind of system, the CA is responsible for the generation, issuance, storage, maintenance, and withdrawal of public key certificates for users, which requires a significant amount of computing and storage resources.

In 1984, an identity-based encryption (IBE) scheme was presented by Shamir [45], which simplified the management of public key certificates in traditional public key infrastructures. The IBE method directly adopts a user’s identity information as the public key. The private key is generated by a trusted party, called the certification authority (CA), for each user. The CA is responsible for authenticating and issuing the corresponding public key certificates. Since then, the IBE method no longer needs the support of the CA, avoiding the establishment and management of public key infrastructure in traditional public key cryptography system. In 2000, Sakai, Ohgishi, and Kasahara suggested that bilinear maps on elliptic curves can be used to design the identity-based cryptography scheme. In 2001, Boneh and Franklin realized the first practical IBE scheme using bilinear maps on elliptic curves and proved that the scheme is resistant to chosen-ciphertext attacks in the random oracle model [46]. Since then, the bilinear maps on elliptic curves have gradually become the main tool of identity-based cryptography scheme.

In applications, the IBE scheme is typically composed of four algorithms [47].

(a) Setup. Select a security parameter k, and get system parameters (params) and the master key. Params include a limited message space M and a limited ciphertext space C, which are open. The master key is private to PKG.

(b) Extract. Input params, the master key, and ID ∈ {0, 1}^*, and get the private key d. ID is an arbitrary sequence as a public key; d is the corresponding private key. The Extract algorithm is used to extract private keys from given public keys.

(c) Encrypt. Input params, ID, and M₀ ∈ M, and get the ciphertext C₀ ∈ C.

(d) Decrypt. Input params, C₀ ∈ C, and the private key d, and obtain M₀ ∈ M.

The above algorithms must be consistent. That is to say, for a given ID, when the private key d is extracted by the Extract algorithm, there is Decrypt (params, C₀, d) = M₀, where M₀ ∈ M and C₀ = Encrypt (params, ID, M₀).

Based on bilinear maps on elliptic curves, we design our IBE scheme in this paper which slightly differs from the Boneh-Franklin cryptosystem but is equivalent in terms of security. It consists of four algorithms as follows.

Initialization. Let k be a security parameter and q be a k-bit prime. Suppose G₁ and G₂ are two cyclic groups of prime order q and ē : G₁ × G₁ → G₂ is an admissible bilinear map with generator P of group G₁. (See [46] for the definition of admissible bilinear maps). Assume that identities are L-bit strings (where L is polynomial in k).

Consider a cryptographic hash function H : {0, 1}^L → G₁. The public key generator (PKG) chooses s ∈ {0, 1, . . . , q − 1} uniformly at random and computes P_{pub} = sP. Here s is the master private key, while all other parameters mentioned above are public.

Private Key Generation. For an identity ID, the private key is x_ID = sQ_ID and the public key is Q_ID = H(ID).

Encryption. To encrypt d ∈ G₂ under identity ID, one can compute E_ID(d, r) = (d · ē(P_{pub}, rQ_ID), rP), where r ∈ {0, 1, . . . , q − 1} is picked uniformly at random.

Decryption. Let C := (C(1), C(2)) be a valid ciphertext under identity ID. Then, C can be decrypted as follows:

\[ D_{x_ID}(C) = \frac{C(1)}{\tilde{e}(x_ID, C(2))} \]

The protocol is homomorphic. That is to say, the following equation is satisfied:

\[ E_{ID}(d_1, d_2, r_1 + r_2) = E_{ID}(d_1, r_1) E_{ID}(d_2, r_2), \]

where the product of the two ciphertexts is defined as taking the product of each component of the ciphertexts.

4. Protocol Design and Analysis

Suppose ID_A is the identity of the preschool child and ID_B is the identity of a volunteer helper (who could be one of the preschool children volunteering to contribute his computational resources). We assume the administrator does not collude with the volunteer helper (if there is a risk of collusion, we can extend this protocol by adding more helpers, which is straightforward).

Let n be the total number of preschool children. Assume that, before the health data are transmitted, each preschool child i has been assigned a unique number R_i ∈ {1, 2, . . . , n}, such that no two preschool children have the same number, that is, for any i ≠ j, R_i ≠ R_j.

The protocol includes two phases: a health data submission phase and a decryption phase.

In the i-th round of the health data submission phase, each preschool child j first compares his own number R_j with i. If R_j = i, then he submits:

\[ D_{i,j} = E_{ID_A + ID_B}(d_j, r_{i,j}), \]

where d_j is his health data and r_{i,j} is picked uniformly at random. If R_j ≠ i, then he submits

\[ D_{i,j} = E_{ID_A + ID_B}(1, r_{i,j}), \]

where r_{i,j} is also picked uniformly at random.

Upon receiving the encryptions in the i-th round of the health data submission phase, the administrator computes

\[ D_i = \prod_{j=1}^{n} D_{i,j}. \]
In the decryption phase, the administrator first forwards all $D_i$ to the helper, who computes
\[
\tilde{d}_i = D_{\text{ID}_A} (D_i)
\]
and returns it to the administrator. Suppose $D_i = (D_i^{(1)}, D_i^{(2)})$. Then, the administrator computes
\[
\tilde{d}_i = D_{\text{ID}_A} (\tilde{d}_i, D_i^{(2)}).
\]

The protocol is correct; that is, assuming all involved parties follow the protocol, then $(\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n)$ is a permutation of $(d_1, d_2, \ldots, d_n)$.

**Theorem 1.** The protocol is correct; that is, assuming all involved parties follow the protocol, then $(\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n)$ is a permutation of $(d_1, d_2, \ldots, d_n)$.

**Proof.** It is easy to see that, assuming all involved parties follow the protocol,
\[
\tilde{d}_i = D_{\text{ID}_A} (\tilde{d}_i, D_i^{(2)}) = \frac{\tilde{d}_i}{\tilde{d}_i^2} = \frac{D_{\text{ID}_A} (D_i)}{\tilde{d}_i^2}.
\]
Let $J(i)$ be the value of $j$ such that $R_j = i$. Hence,
\[
\tilde{d}_i = D_i = \frac{\prod_{j \neq i} D_{ij}}{\tilde{d}_i^2} (1, r_{ij})_{j \neq i} = d_{ji}.
\]

Since $J(i)$ is a permutation on $(1, 2, \ldots, n)$, the result is a permutation of $(d_1, d_2, \ldots, d_n)$.

5. Conclusion

With the development of network technology, more and more data need to be transmitted over the network. Related privacy issues also become a hot research topic. We studied the privacy issue of health data transmitted over the network. For the sake of children’s health, a physical examination of preschool children is needed every year. The data need to be transmitted to the education authorities over the Internet for health analysis after examination. In the process of data collection, it is unnecessary for the education authorities to know the identities of the children. Therefore, we designed a privacy-preserving protocol for health data transmission, which delinks the children’s identities from the examination data. The protocol is composed of three algorithms: Setup, Encrypt, and Decrypt. At last, we proved the correctness of the protocol.

References

[1] R. Agrawal and R. Srikant, “Privacy-preserving data mining,” SIGMOD Record (ACM Special Interest Group on Management of Data), vol. 29, no. 2, pp. 439–450, 2000.
[2] R. Bista and J.-W. Chang, “Privacy-preserving data aggregation protocols for wireless sensor networks: a survey,” Sensors, vol. 10, no. 5, pp. 4577–4601, 2010.
[3] C. -Y. Chow, M. F. Mokbel, and T. He, “A privacy-preserving location monitoring system for wireless sensor networks,” IEEE Transactions on Mobile Computing, vol. 10, pp. 94–107, 2011.
[4] C. -Y. Chow, M. F. Mokbel, and T. He, “A privacy-preserving location monitoring system for wireless sensor networks,” IEEE Transactions on Mobile Computing, vol. 10, pp. 94–107, 2011.
[5] S. Zhong, “Efficient, anonymous, and authenticated conference key setup in cellular wireless networks,” Computer Engineering, vol. 34, no. 5, pp. 357–367, 2008.
[6] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE Transactions on Mobile Computing, vol. 18, no. 2, pp. 582–595, 2010.
[7] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[8] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[9] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[10] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[11] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[12] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[13] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[14] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[15] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[16] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[17] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[18] S. Zhong and F. Wu, “A collusion-resistant routing scheme for noncooperative wireless Ad Hoc networks,” IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 582–595, 2010.
[19] T. Chen and S. Zhong, "Privacy-preserving backpropagation neural network learning," IEEE Transactions on Neural Networks, vol. 20, no. 10, pp. 1554–1564, 2009.
[20] S. Zhong, "Privacy-preserving algorithms for distributed mining of frequent itemsets," Information Sciences, vol. 177, no. 2, pp. 490–503, 2007.
[21] L. Chen, J. J. Yang, and Q. Wang, "Privacy-preserving data publishing for free text Chinese electronic medical records," in Proceedings of the IEEE 36th Annual Computer Software and Applications Conference (COMPSAC ’12), pp. 567–572, 2012.
[22] I. V. Goldberg, "Electronic medical records and patient privacy," The Health Care Manager, vol. 18, no. 3, pp. 63–69, 2000.
[23] B. Zhou, J. Pei, and W. Luk, "Temporary survey on anonymization of customer data," in Proceedings of the 24th IEEE International Conference on Data Engineering (ICDE ’08), pp. 588–599, 2008.
[24] S. Haas, S. Wohlgemuth, I. Echizen, N. Sonehara, and G. Müller, "Aspects of privacy for electronic health records," International Journal of Medical Informatics, vol. 80, no. 2, pp. e26–e31, 2011.
[25] A. R. Miller and C. Tucker, "Privacy protection and technology diffusion: the case of electronic medical records," Management Science, vol. 55, no. 7, pp. 1077–1093, 2009.
[26] Z. Hao, S. Zhong, and N. Yu, "A time-bound ticket-based mutual authentication scheme for cloud computing," International Journal of Computers, Communications and Control, vol. 6, no. 2, pp. 227–235, 2011.
[27] V. Oleshchuk, "Internet of things and privacy preserving technologies," in Proceedings of the 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace and Electronic Systems Technology (VITAE ’09), pp. 336–340, May 2009.
[28] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: a survey of recent developments," ACM Computing Surveys, vol. 42, no. 4, article 14, 2010.
[29] B. Zhou, J. Pei, and W. Luk, "Brief survey on anonymization techniques for privacy preserving publishing of social network data," ACM SIGKDD Explorations Newsletter, vol. 10, pp. 12–22, 2008.
[30] J. Krumm, "A survey of computational location privacy," Personal and Ubiquitous Computing, vol. 13, no. 6, pp. 391–399, 2009.
[31] E. Magkos, "Cryptographic approaches for privacy preservation in location-based services: a survey," International Journal of Information Technologies and Systems Approach, vol. 4, pp. 48–69, 2011.
[32] S. A. Buckovich, H. E. Rippen, and M. J. Rozen, "Driving toward guiding principles: a goal for privacy, confidentiality, and security of health information," Journal of the American Medical Informatics Association, vol. 6, no. 2, pp. 122–133, 1999.
[33] J. A. Diaz, R. A. Griffith, J. J. Ng, S. E. Reinert, P. D. Friedmann, and A. W. Moulton, "Patients’ use of the internet for medical information," Journal of General Internal Medicine, vol. 17, no. 3, pp. 180–185, 2002.
[34] L. J. Damschroder, J. L. Pritts, M. A. Neblo, R. J. Kalarickal, J. W. Creswell, and R. A. Hayward, "Patients, privacy and trust: Patients’ willingness to allow researchers to access their medical records," Social Science and Medicine, vol. 64, no. 1, pp. 223–235, 2007.
[35] S. Zhong, Z. Yang, and R. N. Wright, "Privacy-enhancing k-anonymization of customer data," in Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS’05), pp. 139–147, ACM, June 2005.
[36] R. Agrawal and C. Johnson, "Securing electronic health records without impeding the flow of information," International Journal of Medical Informatics, vol. 76, no. 5-6, pp. 471–479, 2007.
[37] G. Szarvas, R. Farkas, and R. Busa-Fekete, "State-of-the-art anonymization of medical records using an iterative machine learning framework," Journal of the American Medical Informatics Association, vol. 14, no. 5, pp. 574–580, 2007.