7-6 Privacy Preserving Technologies for Personal Data

Ryo NOJIMA, Sachiko KANAMORI, Atsushi WASEDA, Keita EMURA, and Takuya HAYASHI

From November 2014 to March 2016, the members of the Security Fundamentals Laboratory focused on the privacy preserving technologies for personal data. In this paper, our activity is introduced.

1 Introduction

Making use of big data collected from social systems and everyday activities of people to gain new knowledge and develop inventions is now becoming an imperative factor for increasing competitiveness. Data related to the behavior and status of individuals (personal data) is said to have especially high utility value, and an urgent issue now is how to utilize personal data while preserving the privacy of individuals.

Earlier in Japan, big data was utilized without paying much heed to privacy. This resulted in incidents that were harshly criticized by society, and many organizations are hesitating to use big data, as they are not sure how to handle the data from the viewpoint of privacy protection. This could lead to a decline in industrial competitiveness in Japan.

Utilization of big data in governmental growth strategies is one of the mainstays of economic revival, and to remove the barrier to use of personal data by businesses and prevent violation of individual rights and interests, there is a need to develop an environment for achieving the use of data for developing new industries and new services and improving the safety and security of the citizens. Further, a policy for reviewing the system for utilization of personal data was determined, and in September 2013, "Personal Data Study Group" was established under the IT Strategy Headquarters, Cabinet Secretariat. In June 2014, the "Outline of Reform of the Personal Data Utilization System" was published. In March 2015, a system reform bill was decided in the cabinet. In September 2015, the "Act for Partial Revision to the Act on the Protection of Personal Information and Act on the Use of Numbers to Identify a Specific Individual in Administrative Procedures" was promulgated. In this system revision, due to the progress of information and communications technology, and the fact that the subjective views of individuals change over time, the Revision Act only determines the framework, and the specific details are to be defined in accordance with cabinet orders, ministerial orders and ordinances, and government guidelines as well as private sector self-regulations. The issue is to establish a specific method for determining how to actually use personal data from the viewpoint of privacy protection.

The study implemented on this issue was not highlighted in the 3rd Medium-Term Plan. However, considering the trend described above, from November 2014 to March 2016, at the Security Fundamentals Laboratory of the Network Security Research Institute, a workshop was held on privacy related problems, experts were interviewed, and cases were collected. This document reports on those activities.

2 What kinds of information should be handled carefully?

Privacy experts were invited to workshops held 9 times, and working groups held 3 times, in the course of the activities. In the workshops, experts from fields other than security technology (informational ethics, education, law, risk management, and psychology) gave talks. They spoke about how privacy is handled in each of their fields, privacy problems arising in their fields, etc., and opinions were shared.

Among these, pressing issues that drew interest across fields were:

- What kinds of information are considered private information?
- What kinds of points should be heeded when handling information?

The issue of what information is considered privacy information is a major concern socially, but privacy information is generally thought of as:

1. Facts of personal life, or any information from which facts of one's personal life can be construed
2. Information which a private person would not want to be disclosed based on the feelings of the general public, that is, based on the feelings of the general public, information that would cause insecurity or psychological burden if disclosed
3. Information which is still not known to the general public

However, this interpretation is very difficult to work with in engineering. For example, "feelings of the general public" in item 2 is too subjective to be handled in engineering. This activity focused on engineering's handling of "feelings of the general public," and the medium and long term goals were considered as follows:

- Estimate the extent of resistance from people providing information, when doing experiments or when deploying services
- Concept of obtaining consent and automation of obtaining consent

To set these targets, it is important to understand how "feelings of the general public" are subjective, that is, how they change due to conditions. First, to understand how "feelings of the general public" have changed with the times, a questionnaire survey was conducted in 2015, and compared with the 2010 results. Considering the growing popularity of social media, we imagined that resistance to disclosing personal information is declining, but the results of 2015 show that resistance to disclosing self-information is still stronger than expected. Further tests were also conducted to understand how changes in the information collectors, acquisition places, acquisition periods and conditions of information acquisition affected the degree of resistance. Each test was conducted on 2,000 people (males: 1,000, females: 1,000; target ages: 20 to 69, 400 people from each age group; target area: all of Japan) using the questionnaire survey as the basis.

This test led to an unexpected conclusion that reduction in the length of the collection period, not in the amount of collection information or the number of collectors, was most effective in decreasing the resistance against information provision in information providers (for details, see [1]).

The amendment to the Personal Information Protection Act promulgated in September 2015 stipulates the prohibition of information usage for any purpose other than the defined purpose, and restricts provision to third parties without obtaining prior consent of the person in question. The result of this time's survey shows there is an unexpectedly high degree of resistance towards information provision, and hereafter, information collectors may need to reconsider their procedures for obtaining consent. If information collectors, for fear of violating the law, try to obtain consent for any information they collect as they do now, this could rather make the requirement of obtaining consent a mere facade. A method for automating consent acquisition is being studied in the research as one method of meaningful consent acquisition. To implement automation of consent acquisition, we plan to further subdivide the conditions for information collection.

3 Privacy protection technology

With the purpose of collection of information related to privacy protection technology which can be applied for utilizing personal data, existing methods were analyzed and a new method was proposed.

3.1 Privacy protection technology using cryptography: Group signature maintaining anonymity depending on the period, and its application in road-to-vehicle communications

In the reporting system in road-to-vehicle communications of data fetched by cars (traffic jam information, road conditions, temperature, speed, location information, etc.), to prevent mixing of unauthorized data, it is important to verify the legitimacy of the vehicle. However, if the conventional digital signature or message authentication code is used, the vehicle is identified uniquely, so its location information becomes public; for example, residence or workplace information can get leaked. To prevent this, group signatures are employed. A group signature is a signature scheme which can only prove that the person signing on behalf of their group belongs to the group, and many road-to-vehicle communication systems based on group signatures are proposed for protecting the privacy of vehicles [2][3]. However, because of its strong anonymity, there is the problem of lacking helpful information which is acquired from route information of a "signature created by the same vehicle (linked)." Due to its anonym-

ity, another problem cited is that the key revocation process is inefficient when a car is scrapped or a signing key is leaked. On the other hand, since pseudonyms are always linked, it becomes a problem from the viewpoint of privacy. In a Security Credential Management System (SCMS) [4], the pseudonym certificate issuing institution issues timely certificates to vehicles, so that route information, etc. can be acquired while maintaining privacy to a certain extent, and the vehicles to whom the certificates are issued are limited, in order to enable signing key revocation, but a problem is that the cost per vehicle for updating the certificates increases.

This study proposed a group signature with time-token dependent linking [5]. Signatures created in a certain period can be linked, but in different periods, these signatures are guaranteed to be unlinkable in the sense of the group signature. Vehicles are not required to undergo any process such as updating certificates, and the signers are also not required to take any step to revoke signing keys of other signers. This makes it possible to handle the key revocation process efficiently in case a car is scrapped or a signing key is leaked, etc. Figure 1 shows an outline of the proposed method. The signature is created according to the trajectory of the travelling vehicle which is shown in colored circles on the map (near Musashi-Koganei) based on OpenStreetMap [6]. It is guaranteed that the signatures created by a car during a certain period (shown in the same color circles on the map) would represent that they are created by the same vehicle, but would not identify the vehicle itself (linked). Thus, the vehicular path driven in a certain period can be acquired. On the other hand, information on which vehicle created the signatures can never be traced from signatures created in different periods (different colored circles in the map). That is, the linked period is controlled, and route information can be collected while maintaining privacy. The signature generation efficiency also compares favorably with usual signatures (DSA, etc.), and the signature can also be generated for vehicles which do not necessarily have a high computational capacity.

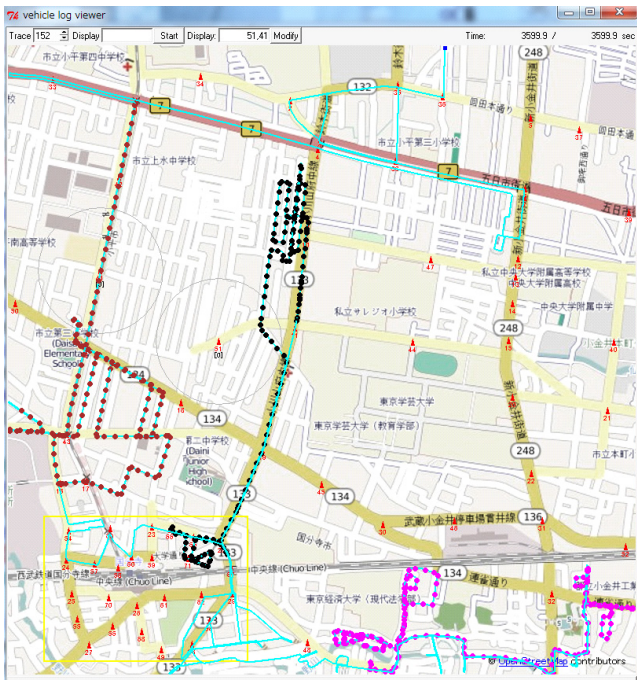

Fig. 1 Route information collection in road-to-vehicle communication system using group signature with time-token dependent linking

3.2 Privacy protection technology without using cryptography: Evaluating the randomized response and its extension by differential privacy

With the growing awareness of privacy in the present day, devices for transmission of information that can identify individuals, and methods of identifying individuals from data collected, are being judged harshly. As a result, the issue of how to secure privacy of transmitted data has become a major concern. One of the privacy protection methods is the randomized response method, proposed by Warner [7]. Information which may interfere with privacy is collected with the help of a questionnaire, through questions like whether the person has an arrest history, in the randomized response method. It is not hard to anticipate that a respondent would hesitate to truthfully answer questions of this kind, even if asked sincerely. The questionnaire is considered to be taken by the following method.

Now, the person asking questions prepares three cards. The first card has a question that the person asking questions originally wants to ask, the second card has "Answer *Yes*" and the third card "Answer *No*" written on them. The person asking questions prepares a box, puts the three cards in the box, and hands it to the respondent. The respondent takes out only one card from the box, reads it, and puts it back in the box without showing it to the person asking questions. Finally, the respondent replies to the person asking questions, according to the content written on that card.

In this way, for example, even if the respondent replies "Yes," it is difficult for the person asking questions to make out whether the respondent is truly answering Yes to the question written on the first card, or just says "Yes" in response to the instruction written on the second card.

(Similarly, even if the respondent replies "No," it could be that the person just says "No" in response to the instruction written on the third card, although he/she should have said "Yes" as his/her true answer to the question written on the first card.). Therefore, this can be expected to reduce resistance to answering questions.

On the other hand, the percentage of respondents answering "Yes" to the original question can be easily estimated by the person asking questions. For example, when the questionnaire is given to a total N number of people, and among them Y number of people replied "Yes." Then, the percentage of people answering "Yes" to the original question can be estimated as 3Y/N-1.

The randomized response method is simple and useful, and hence it has become the foundation of the current topic of anonymization technology. Although it has many applications and many extended methods, it is difficult to find a secure one. In this activity, we evaluate the security of the randomized response method, its extended methods, and the application methods. The differential privacy proposed by Dwork [8] was used in our evaluation. Some of these results are given below.

- The randomized response method mentioned above was generalized, and its differential privacy was derived.
- Negative Survey [9] was evaluated together with its extension (Selective Negative Survey [10]), which showed that they did not fulfill differential privacy.
- Negative Survey fulfilled differential privacy when it was repeated two or more times.

Further, using these results, location information privacy and differential privacy of information fetched from the sensors were also evaluated [11].

4 Summary

Due to the amendment to the Personal Information Protection Act, etc., hereafter, cases related to privacy are likely to increase more than ever. The Security Fundamentals Laboratory will continue this study, with the aim of building mechanisms for using information while protecting the privacy of users.

References

1 S. Kanamori, R. Nojima, H. Sato, N. Tabata, "A study of willingness for private information providing," Symposium on Cryptography and Information Security, 2016.

2 J. Guo, J. P. Baugh, and S. Wang, "A Group Signature Based Secure and Privacy-Preserving Vehicular Communication Framework," in Mobile Networking for Vehicular Environments, 2007, pp.103–108.

3 Q. Wu, J. Domingo-Ferrer, and Ú. González-Nicolás, "Balanced trustworthiness, safety, and privacy in vehicle-to-vehicle communications," IEEE T. Vehicular Technology, vol.59, no.2, pp.559–573, 2010.

4 W. Whyte, A. Weimerskirch, V. Kumar, and T. Hehn, "A security credential management system for V2V communications," in IEEE Vehicular Networking Conference, 2013, pp.1–8.

5 K. Emura and T. Hayashi, "A light-weight group signature scheme with time-token dependent linking," in LightSec, 2015, pp.37–57.

6 OpenStreetMap: https://www.openstreetmap.org/

7 S.L. Warner, "Randomized response: a survey technique for eliminating evasive answer bias", Journal of the American Statistical Association (Taylor & Francis) 60 (309): 63–69, 1965

8 C. Dwork, "Differential Privacy," International Colloquium on Automata, Languages and Programming, 2006.

9 F. Esponda, V.M. Guerrero, "Surveys with negative questions for sensitive items," Statistics & Probability Letters Volume 79, Issue 24, 15, pp.2456–246, 2009.

10 Shunsuke Aoki, Kaoru Sezakim, "Privacy-Preserving Data Mining with Perturbation for Multidimensional Data," Technical Report of IEICE, vol.114, no.65, pp.143-147, 2014.

11 A. Waseda, R. Nojima, "Evaluation for randomized response techniques using differential privacy," Symposium on Cryptography and Information Security, 2016.

Ryo NOJIMA, Ph.D.

Senior Researcher, Security Fundamentals Laboratory, Cybersecurity Research Institute
Cryptography, Cryptographic Protocol

Sachiko KANAMORI

Technical Expert, Security Fundamentals Laboratory, Cybersecurity Research Institute
Privacy

Atsushi WASEDA, Ph.D.

Senior Researcher, Security Fundamentals Laboratory, Cybersecurity Research Institute
Information Security

Keita EMURA, Ph.D.

Senior Researcher, Security Fundamentals Laboratory, Cybersecurity Research Institute
Cryptography

Takuya HAYASHI, Ph.D.

Researcher, Security Fundamentals Laboratory, Cybersecurity Research Institute
Cryptanalysis, Efficient Implementation