Journal of the Korea Convergence Society
Vol. 9. No. 10, pp. 55-59, 2018

ISSN 2233-4890
https://doi.org/10.15207/JKCS.2018.9.10.055

Private information protection method and countermeasures in Big-data environment: Survey

Sunghyuck Hong
Professor, Div. of Information & Communication, Baekseok University

빅데이터 환경에서 개인민감정보 보호 방안 및 대응책: 서베이

홍성혁
백석대학교 정보통신학부 교수

Abstract Big-data, a revolutionary technology in the era of the 4th Industrial Revolution, provides services in various fields such as health, public sector, distribution, marketing, manufacturing, etc. It is very useful technology for marketing analysis and future design through accurate and quick data analysis. It is very likely to develop further. However, the biggest problem when using Big-data is privacy and privacy. When various data are analyzed using Big-data, the tendency of each user can be analyzed, and this information may be sensitive information of an individual and may invade privacy of an individual. Therefore, in this paper, we investigate the necessary measures for Personal private information infringement that may occur when using Personal private information in Big-data environment, and propose necessary Personal private information protection technologies to contribute to protection of Personal private information and privacy.

Key Words : Privacy protection, Big-data analysis, 4th industrial revolution, Big Data Security

요 약 4차 산업혁명 시대에 혁심기술인 빅데이터는 보건, 금융, 유통, 공공부문, 제조업, 마케팅 등 다양한 분야에서 서비스를 제공하고 있으며, 정확하고 신속한 데이터 분석을 통하여 마케팅 분석과 미래 설계에 매우 유용한 기술이며, 앞으로 더 발전할 가능성이 매우 높다. 하지만, 빅데이터 활용 시 가장 큰 문제점이 개인정보 보호와 프라이버시 문제이다. 빅데이터를 이용하여 분석을 통해 기존에 알지 못했던 개인의 취향 및 행동을 분석될 수도 있고, 이러한 정보들은 개인의 민감한 정보이자 개인의 프라이버시 침해가 될 수 있다. 따라서 본 논문에서는 빅데이터 환경에서 개인정보를 활용할 때 발생 가능한 개인정보 침해에 대한 필요 사항들을 분석하여, 그에 따른 필요한 개인정보보호 기술을 제안하여 개인정보 보호와 사생활 보호에 기여하고자 한다.

주제어 : 개인정보 보호, 빅데이터 분석, 4차산업혁명, 빅데이터 보안

1. Introduction

Big-data is no longer an icon of future innovation. However, it is solid positioned as a means of solving the challenges dealing with mankind. Now, it is not a geographical discussion about 'need Big-data?' But it seems that the trend is changing as 'how to use it to create high value'. The use of Big-data and the shield of personal's private information are obviously bilateral. If you emphasize the use of data, your privacy that you

*This research is supported by 2018 Baekseok University research fund.
*Corresponding Author : Sunghyuck Hong (shong@bu.ac.kr)
Received August 10, 2018 Revised August 30, 2018
Accepted October 20, 2018 Published October28, 2018

do not want to disclose will inevitably be violated, and if you emphasize privacy, you will only have difficulty with Big-data research and you may have difficulty in achieving your public purpose. Therefore, the state should take the policy into consideration.

This paper describes the definition and trend of Big-data in Chapter 2, the problems of privacy infringement in Chapter 3, and the protection of private information in Chapter 4. Finally, Chapter 5 concludes the paper.

2. Big-data Trend

2.1 Big-data Trends

2.1.1 Overseas Trends

The UK think tank Economic and Business Research Center estimates that the economic impact of the analytics business on Big-data over the next 5 years will create more than 58,000 new jobs [1-3].

2.1.2 Domestic Trends

If you are interested in Big-data, I think Korea is number one in the world, but despite this interest, the national Big-data market likes not to be formed properly. Currently, large companies that have a large amount of data are in the process of ordering projects.

2.1.3 Example of Privacy Attack

Worker A downloaded a smart phone application that could get discount coupons based on his credit card usage. A who uses an Android phone had to press his consent several times in a pop-up asking whether he would be allowed to access location information, call history, contacts, and SMS from the time the app was laid down. If you do not allow access, you'll be warned that there may be restrictions on your app usage. I remember that some day when I bought a discount coupon from another app to buy something from a large mart, I was told that I could not use it because I disagree with the access permission. The long line got a glimpse of glee without knowing what to do with a smart phone at the front desk. After that, I clicked on the agree button, but I do not know how my information is collected and used [11].

3. Infringement of Personal private information

Big-data has recently become an important infrastructure technology. Among them, CRM (Customer Relationship Management) activities, which use marketing data, can analyze customer's data as well as their behavior, Even marketing services include affiliate marketing, which also uses cost forecast data based on sales performance. Location based service (GPS). Especially, social network service (SNS) is used as a repository for freely expressing individual daily information, and many people provide personal private information such as their school, residence, contact, and marriage status to individual SNS. For example, when an assault by a middle-school student is an issue, people can easily search the portal or search for Personal private information such as student's name, photo, residence, school, etc. This is because unofficial organizations called Nurin find out the Personal private information described in the SNS of the perpetrator and spread it indiscriminately. Personal private information is exposed to other people like this. Users with malicious intent can search data records of others that they want only by mouse clicking. Therefore, the types of privacy infringement that can occur in the Big-data environment are separated step by step [4]. Table 1 shows the various types of misuse for collecting data from unknown users.

Table 1. Step of Privacy Infringement

Type	Action
inapposite Approach and collection	• Without informed consent Collection of Personal private information
inapposite analysis	• Information collected improperly analysis, unspoken history analysis of information

inapposite monitoring	• Individual's Internet without consent • Monitoring activity (cookies)
Inappropriate migration	• Illegal transactions such as transferring Personal private information to a third party
Unwanted sales behavior	• Unauthorized commodity advertising, advertising information transmission behavior
Improper storage	• The external leakage due to the insufficient information security and the act that the Personal private information does not destroy after the purpose of collecting information

3.1 Personal private information infringement cases reported in the news

According to the Personal private information Names Identification Guideline, which was promoted by Park Geun-hye government as the reason for activating Big-data, it was revealed that it provided about 340 million Personal private information to companies [5]. According to the data released on July 9 by Rep. Cho Hyerun of the National Assembly Science and Technology Information Broadcasting Communication Commission, obtained by four non-discriminatory agencies such as Korea Internet Promotion Agency, Korea IT Industry Promotion Agency, Guideline, it has been confirmed that the agency has provided a combination of Personal private information to companies [6].

4. Personal private information Protection Plan

4.1 Non-identification of Personal private information

As Personal private information is exposed, protection must be inevitable. Big-data often contains individual information used by public companies such as banks. Therefore, it is the non - discrimination which is necessary to protect the Personal private information of the individual who owns the Personal private information of the individual. The non-discrimination is made up of a new data set so that the data can not be identified by combining the corresponding collected Personal private information [10][12]. The types of data sets are a pseudonym processing (replacing key identifiers with other values), aggregate processing (summing data or partial aggregation), deleting data values (deleting part or whole), categorization The part or whole of the identifier is not shown) [13].

However, non-discrimination techniques are not always secure because they can be re-identified and information can be found. A Non-identification method has a risk that there is a little bit possibility of re-identification. The information that has not been re-identified is likely to lose value as Big-data analysis and utilization information because of the large data loss. Therefore, strong regulation and management of non-discrimination strength is required to protect Personal private information in Big-data [7].

4.2 Non-Identification Process Processing Architecture

In the processing structure of the non-discrimination process, data is collected from the data subject to which the data refers. The collected personal data is combined into a data set containing Personal private information [8]. Non-discrimination creates a new data set so that it can not identify the data. This dataset can be used internally by the institution instead of using the original dataset to reduce the risk of personal privacy.

Non-discrimination can be performed in the data collection phase (flow ②) or when the identified data is collected but the identification information is not really needed (flow ③). That is, it is not necessary to collect identifiers that are not needed for data management. Instead, non-discrimination can be applied after data conversion and before data storage to avoid obtaining identification information (flow ③). If fully identified data is needed by the organization, the identification information indicates that the data is used for data use Should be deleted before it can be published as a dataset. This dataset can then be provided to the

recipient of the trusted data. To the recipient associated with additional, administrative controls such as data usage agreements. Other words, the data might be made freely available to a number of unknown user, such as releasing unidentified data to the Internet [14][15].

Applying a non-discrimination process based on the data life cycle model can reduce the risk of personal privacy and facilitate the public disclosure process. However, because of the inter-relationships among users involved in the data flow, it is effected by when the non-identification process is to be performed. Figure 1 shows data life cycle. They can be done before data collection (flow chart ①), after data collection (flow chart ②), or before data storage (flow chart ③), or before sharing data with the next participant (flow chart ④).

Fig. 1. Data Life Cycle

4.3 Information Security Management System

The certification body reviews and certifies whether the information protection management system that establishes, manages, and operates the information security management system that is appropriate for a specific organization and maintains, manages, and operates the information security management system in order to protect key information assets. In other words, it is a comprehensive management system including administrative, technical, and physical protection measures to secure the stability and reliability of information and communication network. In the IT era, it aims to cope with the paradigm shift, the cyber infringement risk at all times, and organically manage various security measures.

4.4 Personal private information Management System

To assess the degree of risk of all processes of collecting, using, providing, and destroying Personal private information, documenting the risk management procedures, and taking necessary measures accordingly, in order to ensure that the company has technical, administrative, It is a comprehensive system that continuously operates and manages. If Personal private information is leaked, the benefit of PIMS certification is to reduce penalties and penalties within 50% of the damage done by the victim and grant the same benefits as the ISMS certification [9].

5. Conclusion

Currently, there are no legal provisions in Korea that can provide clear protection. So far, individuals have to write Personal private information on the web or pay close attention to their consent. In addition, similar Personal private information management system (Personal private information Management System) and information security management system (Information Security Management System) to establish the practical use of many companies to gain credibility of information, such as the establishment of Personal private information protection system. It is necessary for the government to abolish the vague current standards for judging whether the government can lead to defects, and to give companies benefits such as exemption from investigation of current status of Personal private information management and penalization discounts on Personal private information disclosure accidents. In addition, the integrated management system law should establish basic rights to overcome the problems and limitations that only the Personal private information protection law itself has, and clearly define the information rights and

obligations of information processors by specifying the Personal private information concept of each individual . In addition, the task force for information human rights protection should actively engage in information security issues and carry out victim relief.

Although there is a problem that the use of Big-data should be prevented from being overly focused on the protection of information in the Big-data environment, efforts should be made to continuously utilize the Big-data.

REFERENCES

[1] W. Fan & A. Bifet. (2013). Mining Big-data. ACM SIGKDD *Explorations Newsletter, 14(2)*, 1.

[2] N. Khan, I. Yaqoob, I. Hashem, Z. Inayat, A. W. Mahmoud, M. Alam, M. Shiraz & A. Gani. (2014). Big-data: Survey, Technologies, Opportunities, and Challenges. *The Scientific World Journal*, 1-18.

[3] S. Hong. (2017). Secure and light IoT protocol (SLIP) for anti-hacking. *Journal of Computer Virology and Hacking Techniques*, *13(4)*, 241-247. doi:10.1007/s11416-017-0295-5.

[4] P. Tallon. (2013). Corporate Governance of Big-data: Perspectives on Value, *Risk, and Cost. Computer*, *46(6),* 32-38.

[5] H. Varian. (2014). Beyond Big-data. *Bus Econ*, *49(1),* 27-31.

[6] S. Hong. (2017). Research on IoT International Strategic Standard Model. *Journal of the Korea Convergence Society*, *8(2)*, 21-26. doi:10.15207/jkcs.2017.8.2.021

[7] S. Hong. (2014). Analysis of DDoS Attack and Countermeasure: Survey. *The Journal of Digital Policy and Management*, *12(1),* 423-429. doi:10.14400/jdpm.2014.12.1.423

[8] S. Hong. (2013). Disconnection of Wireless LAN Attack and Countermeasure. *The Journal of Digital Policy and Management*, *11(12),* 453-458. doi:10.14400/jdpm.2013.11.12.453

[9] S. Hong, S. Lim & J. Song. (2011). Unified Modeling Language based Analysis of Security Attacks in Wireless Sensor Networks: A Survey. *KSII Transactions on Internet and Information Systems*, 805-821. doi:10.3837/tiis.2011.04.010

[10] S. Hong. (2015). Two-channel user authentication by using USB on *Cloud. Journal of Computer Virology and Hacking Techniques*, *12(3)*, 137-143. doi:10.1007/s11416-015-0254-y

[11] R. Smolan & J. Erwitt. (2012). *The human face of Big-data.* Sausalito, Calif: Against All Odds Productions.

[12] S. U. Cha. (2014). Big-data Protection of the environment and privacy. *IT and Law Studies, 8,* 193-259.

[13] H. M. Choi. (2014). [Encouragement Prize] A study on the inconsistency problem of the Big-data era and current privacy laws and their solutions. *IT and Legal Research*, *8*, 357-382.

[14] S. W. Heo. (2014). Big-data Legal Issues in Korea. *Journal of Law & Economic Ragulation, 7(2),* 7-21.

[15] K. J. Park. (2012). Big-data Eco System. *Industrial Engineering Magazine*, *19(3)*, 41-47.

홍 성 혁(Sunghyuck Hong) [종신회원]

・2007년 8월 : Texas Tech University, Computer Science (공학박사)

・2007년 9월 ~ 2012년 2월 : Texas Tech University, Office of International Affairs, Senior Programmer

・2012년 3월 ~ 현재 : 백석대학교 정보통신학부 부교수

・관심분야 : Network Security, Hacking, Secure Sensor Networks

・E-Mail : shong@bu.ac.kr