Int. J. Advance Soft Compu. Appl, Vol. 8, No. 3, December 2016
ISSN 2074-8523

Big Data and Data Protection: Issues with Purpose Limitation Principle

Norjihan Abdul Ghani[1], Suraya Hamid[2], and Nur Izura Udzir[3]

Information System Department
Faculty of Computer Science & Information Technology
[1]University of Malaya
50603 Kuala Lumpur Malaysia
Information System Department
Faculty of Computer Science & Information Technology
[2]University of Malaya
50603 Kuala Lumpur Malaysia
Department of Computer Science
Faculty of Computer Science & Information Technology
[3]University Putra Malaysia
43400 UPM Serdang
Serdang Malaysia
e-mail: norjihan@um.edu.my, Suraya_hamid@um.edu.my,
izura@upm.edu.my

Abstract

In this big data era, more and more personal data have been used and further analyzed. Big data analytics has changed the traditional forms of data analysis and create a new predictive approach to knowledge and investigation. It then arise the issue of privacy and data protection related to personal data. This paper will discuss the personal data protection in big data era which related to purpose specification and limitation principle.

Keywords: *big data, big data analytics, personal data protection, data privacy.*

1 Introduction

Today, recent technological and business developments have given rise to a new understanding of personal information. Therefore, more personal data is being collected, processed and transferred than ever before, especially in this big data era. Big data analytics enabled us to analyze data and turn it into useful or

interesting information so that it can be used for any purpose and interest. It will predict user's need in the future. However, the most important things here is when individuals should be aware of the situation and take necessary measures to protect our own privacy, and most importantly change their perception of privacy.

This paper discussed the main issues related to personal protection in big data. Even though there exist many issues related to data privacy and protection, however, in this paper, will highlight on purpose specification and limitation principle.

2 Big data and Personal data

Personal data is data that relates to an identifiable living individual. 'Identifiable' means that the individual can be identified from that data, either alone or in combination with other information [2]. Personal data means any kind of information (a single piece of information or a set of information) that can personally identify an individual or single them out as an individual [1]. Identifiable means that the individual can be identified from that data, either alone or in a combination with other information [2]. Every personal data being collected and used, organizations must ensure that they are complying with their obligations under data protection act. One key data protection requirement is to ensure that processing of personal data must be fair complied with the act.

Currently, most countries have their own data protection act; including Malaysia which has established MPDPA 2010 with the main purpose is to protect personal data. Most of the data protection acts were established with the guidelines from the OECD Guidelines [6]. In [6], the OECD Guidelines took a comprehensive approach with its specified principles; covering data collection, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability. Purpose limitation is one of the principles which defined that personal data must be obtained for specified, explicit and legitimate purposes and must not be further processed for any other purpose that is compatible with the original purpose.

More and more personal information being used which expresses its increasing value to organizations especially in the big data era. The emergence of big data enables the businesses and organizations to improve the decision making and create new opportunities through new approach of data analysis. Big data has changed the traditional forms of data analysis and create a new predictive approach to knowledge and investigation which known as big data analytics. Big data analytics often has three main characteristics which are different to those of traditional data analysis: use of algorithms, using 'all the data' and repurposing data [13].

Applying big data analytic technique or tools obtain personally identifiable information that can be used for whatever purpose [10]. In [2], even though big data analytics do not involve personal data, there are still many examples of big data analytics that do involve processing personal data, from sources such as social media, loyalty cards and sensors at clinical trials. With big data analytics, individuals will receive an offer of a products or services based on their characteristics, which they refer to personal data such as age, preferences and life style. Thus, findings derived from big data analytics can be applied to marketing to individuals.

3 Issues and Challenges Related to Personal Data Protection

The biggest challenge for big data from a security point of view is the protection of user's privacy [3]. It's because big data frequently contains huge amount of personal identifiable information and therefore privacy of users is a huge concerns It was also highlighted in [4] that users' privacy may be breached under the following circumstances:

i. Personal information when combined with external datasets may lead to the inference of new facts about the users. Those facts may be secretive and not supposed to be revealed to others.

ii. Personal information is sometimes collected and used to add value to business. For example, individual's shopping habits may reveal a lot of personal information.

iii. The sensitive data are stored and processed in a location not secured properly and data leakage may occur during storage and processing phases

The use of big data analytics enable us to identify patterns and trends which may predict people's dispositions, for example related to health, political viewpoints or sexual orientation. The creation of big data therefore permits organizations to create information about data that were never apparent or intended in the source information. This will constitutes information subject to special protection, in fact data controllers must be aware of this risk when compiling and analyzing data [12]. A huge amount of data resource and powerful data analytic techniques make traditional ways of protecting individual's privacy be no longer effective [10]. Individuals have little control over their information's use and disclosure in big data analytics [4]. With data protection, personal data will be able to process if it serves a concrete purpose, permitted by law or with consent of the person in questions. However, it is difficult to verify for big data applications. The purpose limitation principle is not suit anymore in this era of big data. This traditional approach of data protection, are no longer fit for the purposes for which they were designed [11]. Due to this, it is necessary to consider the impact of this new paradigm on the traditional notion of data protection and its regulation [8].

4 Purpose Specification and Limitation Principle in Big Data Era

Purpose specification and limitation is one of the principles exist in any data protection act. The purpose specification principle and the use limitation principle are the traditional pillars of data protection regulations, and, with regard to consumer data protection, the so-called "notice and consent" model (i.e. an informed, freely given and specific consent) represents one of the most used mechanisms to legitimate data processing [15, 16]. According to [1], this principle is a two-step approach that personal data must be obtained for specified, explicit and legitimate purposes (the original purpose) and then must not be further processed for any other purpose (the new purpose) that is incompatible with the original purpose. Thus, any personal data collected must comply with this principle. This principle creates a two-part test: firstly the purpose for which the data is collected must be specified and lawful, and secondly, if the data is further processed for any other purpose, it must not be incompatible with the original purpose.

Big data challenges the principle of purpose limitation, and the principle is a barrier to the development of big data analytics [5]. These critical aspects as stated in [7] concerning the purpose specification limitation have a negative impact on the effectiveness of the "notice and consent" model. For example, big data analytics enable a data analysis using many different algorithms which reveals unexpected correlations that can be used for new purposes. Therefore, the purpose limitation principle restricts an organization's freedom to make these discoveries and innovations. This section will discuss two main issues which involved purpose specification and limitation principle which challenged by big data analytics.

4.1 Repurpose Data

The first issue arises when using data for new purposes. Big data involves reuse of data which it leads to repurpose data. Big data analytics has the ability to uncovering valuable knowledge through compilation of personal data where bigger data sets are putting the principle of purpose limitation under pressure. According to this principle, organizations which use collected personal data as a basis for predictive analysis must ensure that the analysis is compatible with the original purpose for collecting the data [12, 14]. When individuals share data with others, they have a natural expectation about the purposes for which the data will be used. People do not hand over information to a company or the government to do whatever they wish with it. In addition big data analytics also repurposes data that was obtained for a different purpose and in some cases by another organization. In particular, this means considering how the new purpose affects the privacy of the individuals concerned and whether it is within their reasonable expectations that their data could be used in this way. This entails a challenge to

the privacy principle that collected data may not be used for purposes which are *incompatible* with the original purpose for collection.

4.2 New Data from Big Data Analytics

The second issue arises when there is a need to identify the purpose for the new personal data which have been created from big data analytics. In addition to repurpose, big data analytics also has the potential to create a new personal data [2]. Big data can be used to identify general trends, rather than to understand more about individuals or to make decisions about them by using the personal data related to them. Creation of new personal data from big data analytics requires us to provide an approach to get consent from data subject towards their new personal data. With big data analytics, individuals will receive an offer of a products or services based on their characteristics, which they refer to personal data such as age, preferences and life style. Thus, findings derived from big data analytics can be applied to marketing to individuals. Social media data and other data about an individual; could be analyzed to find out about the person's lifestyle in determining their credit rating. In this big data analytics, data collected for one purpose can often be repurposed in ways that greatly benefit society.

5 Conclusion

This paper discussed about personal data protection in big data. Two main issues related with purpose specification and limitation principle have been discussed. Changing the personal data act in not reliable. Due to that, the limits of the traditional data protection require us to outline a new technique in specifying the purpose to make sure the big data analytics fulfil the personal data protection act.

ACKNOWLEDGEMENTS
This research was supported by the Fundamental Research Grant Scheme (FRGS) **(**FRGS/1/2015/ICT04/UM/03/2**).**

References

[1] An Introduction to Data Protection. The EDRI Papers. Issue 06

[2] Big Data and Data Protection. 20 May 2015. Available at https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf

[3] Lafuente, G. (2015). The Big Data Security Challenge. Network Security. 12-14.

[4] Big Privacy: Bridging Big Data and the Personal Data Ecosystem Through Privacy by Design. 20 May 2015. Available at https://www.ipc.on.ca/images/Resources/pbd-big_privacy.pdf

[5] Big Data and Data Protection. 20 May 2015. Available at https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf

[6] Fred Cate, Peter Cullen, and Viktor Mayer-Sch¨onberger. Data protection principles for the 21st century. Oxford Internet Institute, 2013.

[7] Mantelero, A. The Future of Consumer Data Protection in the E.U. Re-thinking the "notice and Consent" Paradigm in the New Era of Predictive Analytics. Computer Law and Security Review. 30. Pp 643-660.

[8] Mantelero, A Giuseppe Vaciago. (2015). Data Protection in a Big data Society. Digital Investigation. Vol 15(2015). 104-109.

[9] Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G., and Guo, S. 2016. Protection of Big Data Privacy. Special Section on Theoretical Foundations for Big Data Applications: Challenges And Opportunities. Vol. 4. P 1821-1834

[10] Qing Tan1and Frederique Pivot. 2015. Big Data Privacy: Changing Perception of Privacy. 2015 IEEE International Conference on Smartcity/SocialCom/SustainCom together with DataCom 2015 ans SC2 2015.

[11] Unlocking the Value of Personal Data: From Collection to Usage. February 2013. World Economic Forum.

[12] Working Paper on Big Data and Privacy. Privacy principles under pressure in the age of Big Data analytics. International Working Group on Data Protection in Telecommunications. 55th Meeting, 5 – 6 May 2014, Skopje

[13] Information Commissioner's Office, "Big data and data protection,"2014.

[14] Big Data and Data Protection. Kemp IT Law, v.1.0, November 2014.

[15] Article 29 Data Protection Working Party. Opinion 03/2013 on purpose limitation. 2013. Available from: http://ec.europa.eu/justice/dataprotection/article-29/documentation/opinion-recommendation/files 2013/wp203_en.pdf [accessed 27.02.14].

[16] Van Alsenoy B, Kosta E, Dumortier J. Privacy notices versus informational self-determination: minding the gap. Int Rev Law, Comp Tech 2014; 28(2):185e203.