Chua, Hui Na; Chang, Younghoon; Wong, Siew Fan; Tan, Chor Min

Conference Paper

Privacy Protection Policy for Big Data Analytics in the Malaysian Telecommunications Sector

26th European Regional Conference of the International Telecommunications Society (ITS): "What Next for European Telecommunications?", Madrid, Spain, 24th-27th June, 2015

Provided in Cooperation with:
International Telecommunications Society (ITS)

Suggested Citation: Chua, Hui Na; Chang, Younghoon; Wong, Siew Fan; Tan, Chor Min (2015) : Privacy Protection Policy for Big Data Analytics in the Malaysian Telecommunications Sector, 26th European Regional Conference of the International Telecommunications Society (ITS): "What Next for European Telecommunications?", Madrid, Spain, 24th-27th June, 2015, International Telecommunications Society (ITS), Calgary

This Version is available at:
https://hdl.handle.net/10419/127131

Privacy Protection Policy for Big Data Analytics in the Malaysian Telecommunications Sector

Hui Na Chua[1], Younghoon Chang[2], Siew Fan Wong[3], Chor Min Tan[4]

[1]Dept. of Computing and Information Systems, Sunway University, Malaysia

E-mail: huinac@sunway.edu.my

[2]Dept. of Computing and Information Systems, Sunway University, Malaysia

E-mail: younghoonc@sunway.edu.my

[3]Dept. of Computing and Information Systems, Sunway University, Malaysia

E-mail: siewfanw@sunway.edu.my

[4]Broadband Traffic Intelligence, Celcom Axiata, Malaysia

E-mail: chormin.tan@celcom.com.my

Abstract

The telecommunications sector has accessed to large amount of data. When use effectively, this Big Data enables the telcos to achieve efficiency and profitability across the entire telecommunications value chain. However, the potential advantage of Big Data may be tempered by increasing privacy concern among users. Countries across different parts of the world including Malaysia have enacted data protection policy. In Malaysia, the Personal Data Protection Act 2010 (PDPA) was officially enforced on November 15th, 2013. To date, its implementation remains challenging and its effect is unclear. This paper attempts to understand the state of data protection policy implementation and its challenges from the viewpoints of three major stakeholders: the users (i.e., the data owners and creators), the telcos (i.e., the data recipients), and the government (i.e., the policy enforcer). Guided by Giddens' Structuration Theory and the Competing Value Framework, semi-structure interview data will be collected from the three stakeholders to understand how differing perspectives of the stakeholders shape the data protection structure/institution and vice versa.

Keywords: Privacy protection policy, big data analytics, telecommunications sector, fair information practices, privacy concerns

Introduction

The telecommunications sector in Malaysia has high potential of benefiting from Big Data. According to a joint-study done by Google and TNS, Malaysia is ranked number one worldwide for smartphone Internet access exclusivity where 35% of the smartphone users use their devices exclusively as the only means of accessing the Internet (Google, 2014, Lee, 2014). In fact, Malaysia along with Singapore, Hong Kong, China and South Korea are the only nations worldwide that use smartphones more than computers as the primary device to access the Internet (Google, 2014; Lee, 2014). The rise of IP networks for telecommunications and mobile applications further reinforces the use of smartphones as the platform for storing and processing information (Qi & Gani, 2012).

The level of smartphone usage connotes the mass amount of data that flows through the telecommunication network every millisecond via user activities such as calls, messages, application download and usage, social networking posts, geo position data, web browsing history, and billing records. These digital footprints are rich data that can be used to create innovative service for end users. In fact, using Big Data, telcos could achieve efficiency and profitability across the entire telecommunications value chain. Therefore, the telecommunications sector, with its ability to collect, store, process, and maintain customer data, are potentially the biggest beneficiary of Big Data (Brown et al., 2011).

However, the ability of telcos to ride on Big Data may be tempered by consumers' privacy concern. While the Malaysia government has enforced the Personal Data Protection Act 2010 (PDPA) in November 2013, its implementation remains challenging and its effect is unclear. This paper attempts to understand the state of PDPA implementation and its challenges from the viewpoints of three major stakeholders: the consumers (i.e., the data owners and creators), the telcos (i.e., the data recipients), and the government (i.e., the policy enforcer). Each stakeholder has differing position that competes within the data protection value system. For example, the data owners want the strictest policy to protect their data, the telcos want more leniency to capitalize on the data, and the government has to establish a policy structure that protects both parties. Through understanding of the differing viewpoints, we believe a more comprehensive and clearer guidelines can be formulated to protect data owners while giving telcos the flexibility to use the data to improve the quality of service and the quality of life.

Literature Review

Big Data

The definitions of Big Data vary greatly. While there is no singularly accepted definition, commonly referred characteristics of Big Data originated from Gartner's report (Laney, 2001) are the "three Vs": large data "Volumes", from a "Variety" of sources, at high "Velocity". The data is captured, stored and analyzed in real-time. Big Data can also refer to the enormous increase in access to data and automated use of information (White, 2012) as well as the enormous amounts of digital data controlled by organizations and authorities (Xu et al., 2011).

Besides the commonly kept structured data in organizations' data warehouses, Big Data can be built on unstructured data gathered from various sources such as social media, communication

texts, multimedia content including images, video messages and sensors (Acker et al., 2013). The complexity and size of data created are far beyond the typical traditional structured databases - not only from the aspect of the data volume, but also the process of capturing, storing, processing, analyzing and deriving insights.

Big Data does not merely involve data alone, it also involves three iterative life cycle (Fayyad, 1996; Hunton & Williams, 2013) phases: (1) data collection, (2) data mining, and (3) analytics application. This is often referred to as Big Data Life Cycle (BDLC).

1. Data Collection

   Big Data requires massive data set in order for its analytics to generate mining algorithms that explore new insights and unanticipated connections among the data analyzed. The massive data set is collected from various sources (PCAST, 2014) including data actively provided by users and passively generated by their devices:

   - Digital data created by users or by a computer surrogate: Examples of data include emails and text messaging, input via mouse clicks, taps, swipes, or keystrokes on a smart device, GPS location data, phone calls details, data associated with most commercial transactions such as credit card swipes, barcode reads, reads of RFID tags, data associated with access key card or ID badge reads and toll road access (remote reads of RFID tags), mobile devices network connection details, and increasingly, data from cars, televisions, and appliances (i.e., the "Internet of Things".) Social media data that are posted would be aggregated might be synthesized in new forms (Clemmitt, 2013).

   - Data from sensors: commonly used sensors are cameras, including videos, which sense visible electromagnetic radiation; and microphones, which sense sound and vibration. Smartphones today contain not only cameras, microphones, and radios but also equipped with sensors for magnetic fields (3D compass), motion (acceleration) and facial recognition.

2. Data Mining

   Data mining, loosely equated to analytics but is only a subset of it. Data mining refers to a computational process that discovers patterns in massive data set (PCAST, 2014). Once a sufficiently massive data set has been collected, it may be analyzed in order to discover correlation among the data that will lead to insights gain in knowledge. Data mining algorithms can be trained to find patterns, each with its own specialized algorithms (Bramer, 2013).

3. Analytics applications

   Big Data algorithms are used to discover piece of knowledge based on data set patterns. Analytics applications consist various computational technologies and algorithms that are used to further create new value. This phase is where an algorithm generated in the data mining phase is applied to produce analytic insights. Depending on the data set input into the algorithm, the insights produced may disclose personal information about the individuals associated with the data (PCAST, 2014).

For the telecommunications sector, Big Data enables the opportunities to achieve efficiency and profitability across the entire telecom value chain (Acker et al., 2013). Massive data from smartphone activities not only creates innovative service for end users, it also enables business opportunities to improve the quality of life. For instance, telcos Orange (for Abidjan, Cote d'Ivoire) and Korea Telecom offered access to anonymized data containing user records of text messages exchanged and local calls to support transport planning in order to reduce traffic congestion and optimize night bus routes (FutureGov, 2014). Similarly, XO communication in the US and Ufone in Pakistan utilized user telecommunications data to analyze the churn rate and identify factors to increase customer retention rate (IBM, 2015).

Smartphone data can also be utilized to analyze migration patterns for managing and monitoring the impacts of local and global socio-economic crises. Researchers are studying migration movements following disasters as a way to understand the spread of infectious disease. For instance, Buckee and her research team (Wesolowski et al., 2013; Buckee et al., 2013) used location data from mobile phones to explore the human moving patterns in Kenya to prevent malaria and other diseases from spreading. Information collected on human travel patterns from mobile phone usage are being utilized to develop predictive models to further fight malaria in the region (Talbot, 2013).

Big Data: Privacy Concern

Big Data creates tremendous opportunity not only for the social economy, but also for different areas ranging from marketing, advertising and credit risk analysis to medical research and urban planning. At the same time, the benefits of Big Data are tempered by privacy concern (Tene & Polonetsky, 2012a). Research showed that users are increasingly concerned about how their personal data are used and their awareness about privacy issues and data protection improves[1].

All data that is created based on user or computer surrogate shares certain characteristics. It is created in identifiable entities for particular purposes. Since they are created by intent, the information that they contain is usually limited for efficiency reasons and good engineering design to support the immediate purpose for which they are collected. When data are created, privacy concerns can arise in two different modes: "over-collection" (obvious) and "data fusion" (subtle).

Over-collection occurs when an engineering design intentionally collects information unrelated to its stated purpose. For example, a user's smartphone could easily photograph and transmit the facial expression to a third party as the user types every keystroke of a text message, or could capture all keystrokes, thereby recording texts that the user has deleted. Data fusion occurs when data from different sources are brought into new and often unexpected phenomena (PCAST, 2014). Individually, each data source may have been designed for a specific, limited purpose. But when multiple sources are processed by techniques of data mining, and the combining of records from diverse sources, new meanings can be found. In particular, data fusion frequently results in the identification of individual people (that is, the association of

[1] European Commission 2011; USC Dornslife/Los Angeles Times 2012.

events with unique personal identities), the creation of data-rich profiles and the tracking of an individual's activities over times. By definition, the privacy challenges from data fusion do not lie in the individual data streams, rather, the challenges are emergent properties of the expanding ability to bring into analytical proximity large, diverse data sets and to process them with new data mining algorithms.

Unexpected information collected from sensors could possibly lead to unanticipated beneficial products and services, but it could also give ways to unanticipated misuse. Today, smartphones with cameras can image a cityscape within several miles (Koonin et al., 2014). Also, the ability exists to sense remotely the pulse of an individual, giving information on health status and emotional state (Durand, 2013). It is foreseeable that sooner or later these capabilities will be present in every smartphone or every wearable communication device.

Sensor data can be combined and mined along with user (or computer surrogate) created data. For example, biometric data is able to provide identity information that enhances the profile of an individual, and data on social network behavior are being used to analyze attitudes or emotions (Feldman, 2013). In brief, more and more information can be gathered and put in a quantified format so it can be tabulated and analyzed (Mayer-Schönberger & Cukier, 2013).

Big Data requires not just data, data mining and analytics algorithms, but also physical platforms such as data centers where the data are stored and analyzed. The related security services used for personal data (PCAST, 2014) are also an essential component of the infrastructure. Once available only to large organizations, the Big Data infrastructure is now available through "the cloud" to small businesses and to individuals. The "cloud" refers not just to the physical hard disk drives that exist in the network with the data, but also to the telecommunications infrastructure of application programs, middleware, networking protocols, and business models that allow that data to be distributed, accessed, and used. This may be detrimental to privacy to the extent that it more effectively hides information exchange from the users (Qi & Gani, 2012).

Personal data regarding individuals' health, location, online activities, network and device used is exposed to scrutiny, raising concerns on profiling, discrimination, exclusion, and loss of control (Solove, 2006). Collected data of personal data if truly anonymised, are no longer considered as personal data, and thus data processing restrictions no longer apply to such data[2]. Commonly, the data used for Big Data analytics is anonymized for the purposes of analysis[3].

However over the past few years, scientific work has repeatedly shown that even anonymized data can often be re-identified to specific individuals (Karr & Reiter 2014; Sweeney et al. 2013; Ohm, 2010; Masiello, 2010; Dwork, 2011; Winkler, 2005; Narayanan & Shmatikov, 2008). On the other hand, there are also shortcomings in the main studies on which re-identification view is based (Cavoukian & Castro, 2014). Nevertheless, it is argued that the issue is not about eliminating the risk of re-identification, but whether it can be mitigated so it is no longer significant.

Another question is the right of organizations to use the customer personal data already in their possession and turn them into anonymized and aggregated data as a commodity sold to others. Examples of such data are location and application data of telecommunications companies

[2] Working Party 29: Opinion 06/2013 on open data and public sector information re-use of 5 June 2013.
[3] Telefonica. Smart Steps. http://dynamicinsights.telefonica.com/488/smart-steps-2. Last accessed on 14.3.2015.

(Steel, 2012). This consequently raises several questions from a privacy perspective, for instance, when can data be considered anonymized? Are organizations allowed to use anonymized and aggregated data without customer consent? Should that consent be granted before use, or is it enough to allow customers to opt out?

Big Data: Privacy Research & Initiative

To tackle Big Data privacy concern, research work have been done in various fields for different applications. Privacy and security issues of Big Data from technological perspectives have been discussed in many literature. For examples, user data privacy through a focus on cloud architecture approach (Cuzzocrea, 2014; Wu & Guo, 2013; Li et al., 2013), privacy aspects of analytics over Big Data (Libaque-Sáenz et al., 2014), security aspects of Big Data query processing surrounding confidentiality and authentication methods (Agrawal et al., 2013; Jang el at., 2014), and machine learning technique for data mining accuracy to satisfy privacy and security requirements (Ishibuchi et al., 2013).

Existing privacy models such as k-anonymity (Sweeney, 2002), l-diversity (Machanavajjhala, 2007), t-closeness (Li & Li, 2007) and the Differential Privacy model (Dwork, 2008) follow a database-centric approach. Privacy models using different approaches such as pay-by-data model (Wu & Guo, 2013), privacy aware query (Jang et al., 2014) and fuzzy genetics based approach (Ishibuchi et al., 2013) are also present. Studies have also been conducted to look at major Big Data privacy and security issues and its policy implication (Kisdi, 2014; Cuzzocrea, 2014). More explanation on existing privacy models has been discussed in Backes et al. (2015).

Recently, developed countries such as Japan, UK, USA, and South Korea initiate specific guidelines for Big Data personal information protection via standard committees[4] and working groups (KISDI, 2014). In 2012, the Cloud Security Alliance established a Big Data working group to identify scalable techniques for data-centric security and privacy problems. The World Wide Web Consortium (W3C) has created several community groups on different aspects of Big Data. In 2013, The United States National Institute of Standards and Technology (NIST) launched its Big Data public working group to support secure and effective adoption of Big Data by developing consensus on the definitions, taxonomies, secure reference architectures and a technology road map for Big Data analytic techniques and technology infrastructure.

The UK Anonymiszation Network[5] plays a role in providing expert advice on Big Data anonymization techniques. As for Malaysia, even though it is at the forefront of the Internet use via smartphones and the government has officially enforced the Personal Data Protection Act 2010 (PDPA) on November 15th, 2013, its implementation of Big Data privacy protection remains challenging and its effect is unclear[6].

Privacy Protection Policy for Telecommunications Sector

[4] ISO/IEC JTC1's data management and interchange standards committee initiated a study on next-generation analytics and big data.
[5] UK Anonymisation Network website http://ukanon.net/ Accessed 25 June 2014
[6] Malaysia Data Security Forum, Kuala Lumpur. 5th February 2015. http://asli.com.my/uploads/20150106162852_Brochure-Data%20Security-v1.pdf

Organizations face challenges revolving around data privacy considerations. For instance, specific details of an individual's buying habits and lifestyle preferences are captured and analyzed through the organizations' websites or by monitoring the social media. These details are all collected without individuals' absolute consent, leading to significant reservations about Big Data. Privacy concerns fortify the demand for tighter regulatory control (El-Darwiche et al., 2014). Privacy and data protection regulations are premised on individual control over information and on principles such as data minimization and purpose limitation. Yet, it is unclear that minimizing information collection is always a practical approach to data privacy (Tene & Polonetsky, 2012a). The policies of privacy and data protection should be balanced (Polonetsky & Tene, 2013) and should not supress the innovation that Big Data can deliver, or its values to the economic and social benefits.

Though the telecommunications sector with its ability to collect and process massive customer data are bound to be the biggest beneficiary of the Big Data trend (Brown et al., 2011) and the IDC forecast also revealed that the telecommunications sector tends to spend extensively more on Big Data projects than other industries (Vesset et al., 2012), there is no guarantee that the potential of Big Data will be fully exploited due to obstacles lie in the way (Beardsley et al., 2014). One of the main obstacles is data privacy challenges.

The current data leverage among Telcos is limited by user data privacy and protection as well as security challenges. The consequences of privacy in the Big Data era are not fully understood and the policies are under-developed (Kshetri, 2014). The immediate task now is to understand the state of the data privacy protection implementation before the telecommunications sector can move forward to leverage on the massive user data. This understanding is novel and essential because policy is the core facilitating factor to promote better personal information protection (Xu et al., 2011).

It is also suggested that policymakers should address the role of consent in the privacy framework (Tene & Polenetsky, 2012b) because presently there are too many processing activities premised on individual consent. Yet individuals are ill-equipped to make responsible decisions about their personal data given (Tene & Polonetsky, 2012a; Brandimarte, 2013; Lundblad & Masiello, 2010). Though there is demand for regulatory control, studies however revealed that trust depends on individual's views on privacy, and these views change rapidly (Nissenbaum, 2011) because it is increasingly difficult for many people to understand where the old norms end and new ones begin. Yet, there is evidence that individuals do not require complete protection, and will willingly share personal information provided that certain social norms are met (Dinev & Hart, 2006). The three factors found that affect these norms: actors (the information senders and recipients); attributes (the types of information about the providers); and transmission principles (the information flows constraints). Figure 1 elaborates the relationships between Telcos, the users and the policymakers (i.e., regulators) by considering the three factors introduced in Nissenbaum (2011).

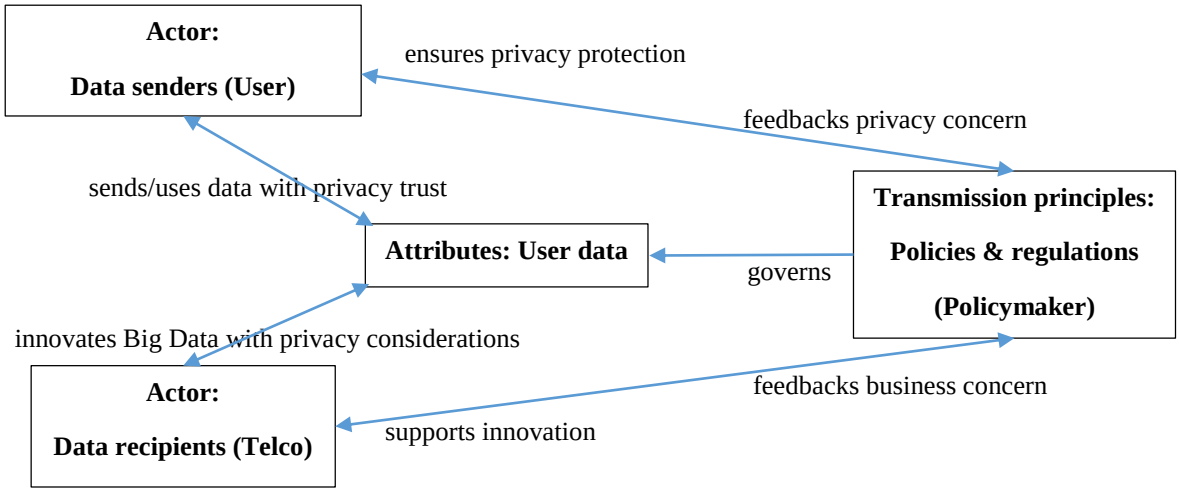

Figure 1: The relationships between Telcos, users and policymakers in privacy protection policy

To maximize the benefits of Big Data and to build trust, policymakers (regulators) need to build a transparent legal framework that supports the telecommunications sector's (also applicable to other industries) growth and innovativeness in service provisioning, rather than hinders it with unnecessary legal burdens. At the same time, users' privacy protection should be insured in regulations. The responsibility is not just on policymakers to establish an environment of trust for users, the industries also have a key role to play by understanding user concerns and regulations to help determine the areas of risk they need to tackle and strengthen to develop their Big Data strategy.

Industries must also work with policymakers because privacy and data protection regulation is constantly evolving. This means that organizations will need to establish a close relationship with national policymakers to ensure compliance and to ascertain that policymakers understand the business concerns at hand and the benefits of Big Data. The relationship between actors is bi-directional since data protection regulations are evolving not only in an attempt to keep pace with technological developments and new approaches of collecting, consuming and sharing personal data, but also to keep pace with attitudes toward privacy (WEF, 2014).

Big Data Privacy Policy: Why the Malaysian Telecommunications Sector?

In this proposed work, we would like to draw attention to Big Data privacy policy in the Malaysia context, particularly on its telecommunications sector. The rationale behind this attention is stimulated by the following observations:

- *Telecommunication sector has the most infrastructure capability for Big Data*

  The telecommunication sector has the most infrastructure capability to collect and operate massive customer data and bound to be the biggest beneficiary of Big Data trend compared

to other industries (Brown et al., 2011). IDC forecast (Vesset et al., 2012) showed that the telecommunications sector tends to spend extensively more on Big Data projects than other industries. Yet, the potential of Big Data has not been fully exploited due to data privacy challenges as one of the main obstacles (Beardsley et al., 2014). By having the Telcos as the leading sector in Big Data trend, the establishment of national telecommunications sector policy can be used as a cross-referenced guidelines for other industries.

- *Malaysia's privacy regulations are unclear though with highest NRI[7] in developing countries*

  Based on the Global Information Technology Report (WEF, 2014), Malaysia establishes its leadership as "the highest ranked economy in developing Asia, and maintains relatively competitive regulatory (25^{th}) and business (24^{th}) environments, and its government continues to use ICTs extensively (9^{th})". Its laws relating to ICTs (13^{th}) highlight "the high priority of this sector in the government's agenda. Business usage (27^{th}) is also strong as firms invest to adopt new technologies and make the effort to become increasingly innovative".

  Though Malaysia is establishing its leadership in developing countries, and is at the forefront of smartphone online use, its implementation of Big Data privacy protection remains challenging and its effort in the rapidly evolving Big Data analytics technology lack clarity[8]. The study of Big Data privacy and its implications from the Malaysia perspective not only helps to understand the challenges nationally, it can also be the basis of recommendations for other developing countries in the ASEAN or Southeast Asia region that shares some commonalities in economy and culture.

- *Different countries with different regulations for different cultural factors*

  Legal inconsistency between countries can obstruct the adoption of Big Data on an international scale. This problem arises, for example, when data are owned by a European Union country, but servers are hosted in the United States. Then the question arises on which privacy law applies? To tackle this problem, there are several existing frameworks for privacy protection (WEF, 2014).

  The Asia-Pacific Economic Cooperation (APEC) has developed a self-regulatory framework in 2004 to set up the principles to ensure a common, minimum level of data protection in order to ease the data transfer among countries where the level of data protection regulation varies broadly. Regulators in the European Union and the United States have a framework to enable data transfer between the two regions without further approval from EU-based regulators. The US and Asian regulators are collaborating around the APEC framework.

  Though with these efforts to set out privacy frameworks, there is still no binding agreement to harmonize regulation around data privacy because different countries have legitimate differences on privacy issues. Cultural factors have a strong bearing on the decision about

[7] Networked Readiness Index (NRI) is used to identify the driving factors and impacts of networked readiness and ICT leveraging, and highlight the joint responsibility of all social actors – individuals, businesses, and governments. Source from the Global information Technology Report, World Economic Forum, 2014.

[8] Malaysia Data Security Forum, Kuala Lumpur. 5th February 2015. http://asli.com.my/uploads/20150106162852_Brochure-Data%20Security-v1.pdf

the right level of data privacy, resulting in a regulatory establishment appropriate for individuals and organizations in a given country (WEF, 2014). Hence, it is crucial to explore the challenges surrounding privacy issues, regulation and implication from the context of Malaysia - a culturally unique country.

Theoretical Framework

Giddens' (1976) Structuration Theory (ST) and the Competing Value Framework (CVF) (Quinn and Rohrbaugh, 1983) will serve as the theoretical foundations for our research. ST analyzes the interplay between the structure and the agents to understand how both dimensions influence each other. When applying to the context of the telecommunications sector, the structure is the policies and regulations enacted by the government while the agents are stakeholders such as the telcos and the users who function within the structure, as well as the government that enacts the policy. The duality concept in ST suggests that the structure imposes certain limitations on how the agents will act but the agents through time could alter the structure to adapt to new changes. Since the structure and the agents (even among the agents) have competing interests, CVF will help to dissect how different interests conflict each other.

Research Methodology

We will adopt a semi-structured interview research method to examine the perspective of the stakeholders involved in data exchange within the telecommunication sector in Malaysia. The stakeholders are (1) the government (being the policy makers), the five largest telcos in Malaysia (being the data recipients and holders), and the end users (being the primary data owners). We will explore how each stakeholders view the state of data privacy protection policy implementation and the challenges faced in an attempt to understand their expectation of privacy protection. The AST and the CVF will frame our interview questions.

Conclusion

Our work presented in this paper is aimed to understand the state of data protection policy implementation and its challenges, and develop a research framework for the telecommunications industry. The framework will provide an understanding of the ecosystem which demonstrates how regulatory policies, users and telcos influence each other and how each stakeholder's interest is ensured or concern is tackled, under the constant evolving of Big Data technology and regulatory policies environment.

This paper presents our research in progress, and the preliminary study on big data privacy concern in telecommunications industry. We will continue this work by conducting in-depth interview with the ecosystem stakeholders to gain understanding of how differing stakeholders' perspectives in shaping the data protection structure and vice versa.

References

Acker O., Blockus, A., Pötscher, F. (2013). Benefiting from Big Data: A New Approach for the Telecom Sector. Strategy & Analysis (formerly Booz & Company).

Agrawal, D., Abbadi, A. E., & Wang, S. (2013, April). Secure and privacy-preserving database services in the cloud. In Data Engineering (ICDE), 2013 IEEE 29th International Conference on (pp. 1268-1271). IEEE.

Backes, M., Berrang, P. and Manoharan, P. (2015). How well do you blend into the crowd?-d-convergence: A novel paradigm for quantifying privacy in the age of Big-Data. arXiv preprint arXiv:1502.03346. Retrieved from http://arxiv.org/pdf/1502.03346.pdf.

Beardsley, S., Enriquez, L., Grijpink, F., Sandoval, S., Spittaels, S., & Strandell-Jansson, M. (2014). Building Trust: The Role of Regulation in Unlocking the Value of Big Data, the Global Information Technology Report 2014, World Economic Forum.

Clemmitt M. (2013). Social Media Explosion: Do social networking sites threaten privacy rights? CQ Researcher, 23 (4), 84-104.

Bramer, M., (2013). Principles of Data Mining. Springer.

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced confidences privacy and the control paradox. Social Psychological and Personality Science, 4(3), 340-347.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. McKinsey Quarterly, 4, 24-35.

Buckee CO, Wesolowski A, Eagle NN, Hansen E, Snow RW, 2013. Mobile phones and malaria: modeling human and parasite travel. Travel Med Infect Dis. 11 (1): 15-22.

Cavoukian, A., and Daniel C. (2014). Big Data and Innovation, Setting the Record Straight: De-identification Does Work. The Information Technology and Innovation Foundation.

Cuzzocrea A. (2014). Privacy and Security of Big Data: Current Challenges and Future Research Perspectives. Proc. of PSBD.

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. Information Systems Research, 17(1), 61-80.

Durand, Fredo, et al. (2013). MIT Computer Program Reveals Invisible Motion in Video. The New York Times, video, February 27, 2013. Retrieved from https://www.youtube.com/watch?v=3rWycBEHn3s.

Dwork, C. (2008). Differential Privacy: A Survey of Results. In Proceedings of the 5th International Conference on Theory and Applications of Models of Computation. 1–19.

Dwork, C. (2011). A Firm Foundation for Private Data Analysis. Communications of the ACM, 54(1), 86-95.

El-Darwiche B.,Koch V.,David Meer D., Shehadi R. T., Tohme W (2014). Big Data Maturity: An Action Plan for Policymakers and Executives. The Global Information Technology Report 2014.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. AI Magazine. Fall 1996. Retrieved from http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf.

Feldman, Ronen (2013). Techniques and Applications for Sentiment Analysis. Communications of the ACM, 56(4), 82-89.

FutureGov (2014). How Seoul uses analytics to find late night bus routes. Retrieved from http://www.futuregov.asia/articles/5340-how-seoul-uses-analytics-to-find-late-night-bus-routes

Giddens, A. (1976). New Rules of Sociological Method. London: Hutchinson.

Google (2014). Consumer Barometer. Retrieved from https://www.consumerbarometer.com/en/insights/?countryCode=MY

Hunton & Williams LLP (2013). Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance. Centre for Information Policy Leadership. February 2013. Retrieved from http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf.

IBM (2015). Creating smarter campaigns with the IBM Advanced Analytics platform. Retrieved from http://public.dhe.ibm.com/common/ssi/ecm/en/imw14738usen/IMW14738USEN.PDF. Last accessed on 13.3.2015.

Ishibuchi, H., Yamane, M., & Nojima, Y. (2013). Learning from multiple data sets with different missing attributes and privacy policies: Parallel distributed fuzzy genetics-based machine learning approach. In Big Data, 2013 IEEE International Conference on (pp. 63-70). IEEE.

Jang, M., Yoon, M., Chang, J.-W. (2014). A Privacy-Aware Query Authentication Index for Database Outsourcing. Proc. of BigComp 2014.

Karr, Alan F. And Jerome P. Reiter. (2014). Using Statistics to Protect Privacy. pp. 276-295 in Privacy, Big Data, and the Public Good: Frameworks for Engagement, edited by J. Lane, V. Stodden, S. Bender, and H. Nissenbaum. Cambridge: Cambridge University Press.

KISDI (2014). Big Data 2.0: Major Issues and Policy Implications. Korea Information Society Development Institute (KISDI) Premium Report 14-10, 1-14.

Koonin, S.E., Gregory D., and Jonathan S. W. (2014). Urban Physics. American Physical Society News, March, 2014. Retrieved from http://www.aps.org/publications/apsnews/201403/urban.cfm

Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. Telecommunications Policy, 38(11), 1134-1145.

Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note, February, 6. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Lee, R.J. (2014). Malaysia leads the world in smartphone usage. 21 November 2014, http://www.businesscircle.com.my/malaysia-leads-world-smartphone-usage/.

Li, N. and Li, T. (2007). t-closeness: Privacy beyond k-anonymity and -diversity. In Proceedings of the 23rd International Conference on Data Engineering (ICDE), 2007.

Li, M., Yu, S., Zheng, Y., Ren, K., & Lou, W. (2013). Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. Parallel and Distributed Systems, IEEE Transactions on, 24(1), 131-143.

Libaque-Sáenz, C. F., Wong, S. F., Chang, Y., Ha, Y. W., & Park, M. C. (2014). Understanding antecedents to perceived information risks An empirical study of the Korean telecommunications market. Information Development, 0266666913516884.

Lundblad, N., and Masiello, B. (2010). Opt-in dystopias. SCRIPTed 7.1, 155-165.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). L-Diversity: Privacy Beyond K-anonymity. ACM Trans. Knowl. Discov. Data, 1(1), 2007.

Masiello, B. & Whitten, A. (2010), Engineering Privacy in an Age of Information Abundance, 2010 AAAI Spring Symp. Series 119, 122. Retrieved from http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/viewFile/1188/1497.

Mayer-Schönberger, Viktor and Kenneth Cukier (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think, Boston, NY: Houghton Mifflin Harcourt, 2013.

Narayanan, A. & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets, 2008 Proc. of IEEE Symp. on Security & Privacy.

Nissenbaum, Helen. (2011). A Contextual Approach to Privacy Online. Daedalus, 140(4), 32-48.

Ohm, P. (2010), Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review, 57(6), 1701-1818.

PCAST. (2014). Big data and privacy: A technological Perspective. Big Data and Privacy Working Group Report. President's Council of Advisors on Science and Technology. May 2014.

Polonetsky, J. and O. Tene. (2013). Privacy and Big Data: Making Ends Meet. Stanford Law Review, 25. Retrieved from http://www.stanfordlawreview.org/sites/default/files/online/topics/PolonetskyTene.pdf.

Qi, H. and A. Gani, (2012). Research on mobile cloud computing: Review, trend and perspectives. Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference on.

Quinn, R. E., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. Management science, 29(3), 363-377.

Solove, D. J. (2006). A taxonomy of privacy. University of Pennsylvania law review, 477-564.

Steel, E. (2012). Datalogix Leads Path in Online Tracking. The Financial Times, September 23. Retrieved from http://www. ft.com/intl/cms/s/2/8b9faecc-0584-11e2-9ebd-00144feabdc0. html#axzz2idgoMkIT.

Sweeney, L. (2002) K-Anonymity: A Model for Protecting Privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst., 10(5), 557–570.

Sweeney, L., Abu, A., and Winn, J. (2013). Identifying Participants in the Personal Genome Project by Name. Harvard University. Data Privacy Lab. White Paper 1021-1, April 24. Retrieved from http:// dataprivacylab.org/projects/pgp/.

Talbot, D., (2013). Big Data from Cheap Phones. MIT Technology Review.com, April 23, 2013. Retrieved from http://m.technologyreview.com/featuredstory/513721/big-data-from-cheap- phones/.

Tene, O., and Polonetsky J. (2012a), Privacy in the Age of Big Data: A Time for Big Decisions, 64 Stan. L. Rev. Online, 63.

Tene, O., and Polenetsky J. (2012b) To Track or Do Not Track: Advancing Transparency and Individual Control in Online Behavioral Advertising. Minn. JL Sci. & Tech. 13, 281.

Vesset, D., Morris, H.D., Little, G., Borovick, L., Feldman, S., Eastwood, M., Woo, B., Villars, R.L., Bozman, J.S., Olofson, C.W., Conway, S., Yezhkova, N (2012). IDC market analysis: Worldwide big data technology and services, IDC #233485, March 2012.

WEF (2014). Global Information Technology Report, World Economic Forum, 2014.

Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., & Buckee, C. O. (2013). The impact of biases in mobile phone ownership on estimates of human mobility. Journal of the Royal Society Interface, 10(81).

White, C. (2012). Big data is the term for a collection of data sets so large and complex that it be- comes difficult to process using on-hand databases management tools or traditional data processing applications.

Winkler, William E. (2005). Re-Identification Methods for Evaluating the Confidentiality of Analytically Valid Microdata. Research Report Series, Statistics #2005-09. U.S. Census Bureau.

Wu, C., and Guo, Y. (2013). Enhanced User Data Privacy with Pay-By-Data Model. Proc. of BigData Conference 2013.

Xu, H., Dinev, T., Smith, J., & Hart, P. (2011). Information privacy concerns: linking individual perceptions with institutional privacy assurances. Journal of the Association for Information Systems. 12(12), 798-824.