A Proposal of Personal Information Trading Platform (PIT):

A Fair Trading between Personal Information and Incentives

Ake Osothongs
School of Multidisciplinary, Department of Informatics
The Graduate University for Advanced Studies
Tokyo, Japan
ake@nii.ac.jp

Noboru Sonehara
Information and Society Research Division
National Institute of Informatics
School of Multidisciplinary, Department of Informatics
The Graduate University for Advanced Studies
Tokyo, Japan
sonehara@nii.ac.jp

Abstract— **Personal information is a valuable asset for service providers. They need personal information for several reasons such as target advertising and personalization. Many methods are used to collect personal information from consumers. However, sometimes more personal information than is needed is collected. Since consumers are now more concerned about their privacy, service providers are finding it a lot more difficult to collect consumer's personal information, which results in less benefit for both consumers and service providers. In this study, we propose a new personal information trading platform that provides benefit to both service providers and consumers. We present the platform design, functions and benefits, a proposed trading model, and the balanced personal information tradeoff model, which is used to measure the balance between the incentive to the service provider and the personal information of the consumer.**

Keywords-personal information trading; privacy; personal information trading; personal information protection

I. INTRODUCTION

In the era of Big Data, people connect themselves to digital networks. Internet users frequently disclose their personal information over online networks. The service providers collect personal information and use it for several purposes, such as for recommendation systems and personalized marketing. Service providers can now easily collect personal information using registration forms and by collecting automatically generated information, such as web access history, location, and sensor information. This growth of information technology raises serious concerns regarding privacy issues for the consumer.

Individuals have their own right to disclose their personal information or not. However, people who lack awareness of the intricacies of information technology are confronted with many risks, such as financial fraud and identity theft, when disclosing their personal information. Although personal information collection can be a serious risk, it creates new opportunities for service providers. Thus, many service providers convince consumers to provide their personal information. The disclosure of personal information is becoming a common online activity. In order to consume these services, the consumers have no choice but to accept to disclose their personal information without any negotiations.

Some authors suggest that the personal information can be a kind of commodity [1], [2]. Personal information becomes a resource that can be used within a company or may sold to others [3]. However, it is difficult to accept being treated as an asset because it is difficult to estimate the value of personal information. There are researchers who have worked on the value of personal information and the results vary. For example, researchers developed a tool called "Cloudsweeper", which aims at finding the value of an email account. The email account value is calculated from the service accounts value that are associated with each email [4]. Otsuki and Sonehara estimated the value of personal information using a SNS utility. The results show an estimation value of the personal information that is based on the cost of the protection of personal information [5]. These researches are just a couple of examples of different opinions about the value of personal information. Even if the value of personal information is difficult to estimate, the service provider still offers the incentives as a reward to the consumer in order to trade their personal information. These rewards possibly affect the self-disclosure decision [6].

Moreover, the privacy and personal identity protection regulations are now very strict in many countries due to the rapid development of information technology. This may create a new problem in the future as business industries might not be able to collect personal information from their consumers or cannot use their already collected personal information. They have to find better ways to attract consumers and get them to provide their personal information and also to accept the usage of it.

This study proposes a personal information trading platform (PIT) and a model for personal information trading. This platform aims at being a medium between the service and identity providers in order to provide maximum benefit for the service provider and privacy protection for the identity provider. In Section 2, we discuss the definition of personal

information that will be used in this research. In Section 3, we discuss the current problems involved in personal information trading and collection, which leads to the design of PIT. In Section 4, we present the design architecture of PIT, its benefits, and propose a trading decision. We discuss an evaluation model of the platform. How to find the maximum benefit for the service provider as well as provide privacy protection to the identity provider is our focus, so we present a case study for this in Section 5. We conclude the future institutional requirements and the challenges for implementing PIT in the final section.

II. PRELIMINARY DEFINITION

A collection of personal information problems usually starts with a question, what is personal information? The terms "personal information (PI)" and "personal data" are widely used interchangeably and have the same meaning. Their definitions have been debated for a long time, especially in the legal terms. However, the definition is still unclear [5], [7].

In early 2000s, most personal information was still retained in offline databases. Personal information problems were not a big issue because it was easy to control. Privacy concerns arose when an individual system stored personal information that could identify an individual such as in a health care system. Many countries back then defined personal information in the same way, and it was usually defined as information that can be identified, directly or indirectly, to an individual or a single person [7], [8]. The term "personally identifiable information" (PII) [9], [10] had the same meaning in many counties.

Information technology have recently become tightly integrated with human life. The definition of personal information has changed to reflect the new age, Big Data, which at small information prices can lead to identifying an individual. Even if a de-identification method such as anonymization, encryption and data shading have been used to protect an individual, researchers have frequently proven that anonymized information can be re-identified to an individual [11], [12].

Therefore, personal information is any information relating to an individual. It can be any information directly or indirectly collected from an individual regardless of its source.

III. PROBLEMS OF PI COLLECTION AND TRADING

We focused on the following personal information collection and trading problems in our study.

Illegal collectors: Even though privacy laws and regulations that deal with the collection of PI have been published in many countries and are widely debated, illegal collectors are still a big problem. Today, there are a lot of untrustworthy PI collectors, such as unknown application providers who ask for consumer's PI when they install an application.

Lack of fair trading: People usually focus on the protection of privacy for consumers but the trading benefits for the service providers usually are ignored. Some researchers suggest that the consumer hide their PI. Alastair et al. introduced MockDroid, a modified version of the Android operating system, which provides a method to return a valid but incorrect information to the service provider [13]. Georgios, Michalis, and Evangelos implemented a SudoWeb module, an extension for the Google web browser, in which the user can select an identity from two prepared identities when using a social login [14]. These are examples of the customer protection that does not return any value back to the service provider.

No opt-out and limit of usage time: The European Commission proposed a new "right to be forgotten" law that allows people to opt-out from the service providers [15]. Nowadays, many websites and applications state their privacy agreements and show the opt-out option. However, sometimes people cannot control the opt-out request. For example, consumers' emails were illegally collected by web crawlers and illegally sold online in the black market. Another problem occurs when PI has been collected and there are no statements as to the limit of usage time of the PI.

Over trading: The service provider always requests as much PI as they desire. They can request more information than is necessary. One problem is it is difficult to find the balance between the protection of privacy and the utilization of the information [16]. We are always faced with this kind of request in many services such as a request on a mobile phone application and social network login. Felt et al. found that the popular Facebook applications require too much PI when a consumer requests their services [17].

Fake PI: The service provider can collect the automatically generated information. However, the PI that is directly collected from the consumer provider is still necessary. Identity providers usually submit fake PI for several reasons. Some criminals use it for criminal activities such as identity fraud [18]. Some create fake profiles to hide themselves when they use the online services such as social network service [19]. Moreover, some professional advisers advise people to use fake information when they do not trust the service provider [20]. This fake PI can be a method to hide their identity on the Internet, but we argue that this leads to a new problem when the legal service providers use the PI for their legal purposes.

IV. PROPOSAL OF PIT: A PERSONAL INFORMATION TRADING PLATFORM

Some reviews for resolving the problems with PI collection and PI trading were described in Section 3. Therefore, we propose a new Personal Information Trading platform (PIT) to solve these issues.

A. Platform design

The proposed PIT in this study is expected to contain the following three components.

a) A Service Provider(SP): provides any services or products that require PI. For example, hotel, travel agencies, and finance institutions, etc.

b) An Identity Provider(IdP): is a consumer and their personal information.

c) An Personal Information Trading Platform (PIT): addresses between a SP and an IdP.

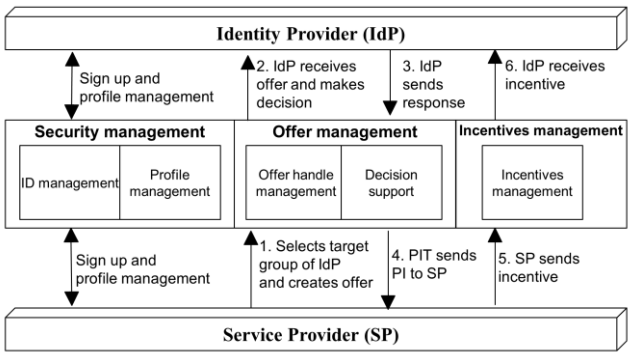

Fig 1. Platform Architecture of PIT

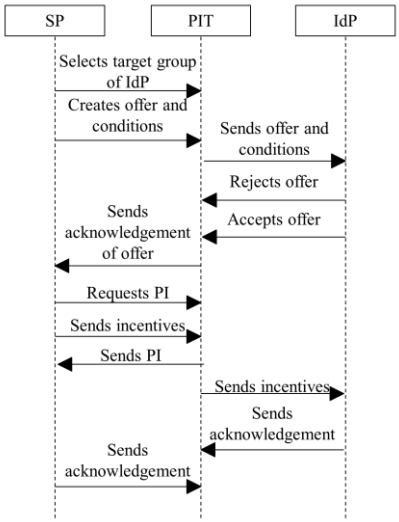

Fig 2. The design of sequence diagram for PIT

Figure 1 shows the platform architecture of PIT, which consists of 3 modules: security management, offer management, and incentive management systems. Figure 2 represents the design of a sequence diagram for PIT. The following activities occur during each trading activity.

1) SP selects an IdP target group and sends an offer.
2) IdP receives an offer and makes a decision.
3) IdP sends a response of the offer.
4) PIT sends the response with the PI to the SP.
5) SP sends an incentive to PIT with the PI request.
6) PIT sends incentives to the IdP.

B. Functions and benefits

SPs currently offer incentives to attract IdPs to provide their PI. The incentive can be monetary such as a gift voucher, product discount, and coupon. However, it is still difficult to collect reliable information from an IdP as described in Section 3. The PIT is a platform that will be placed between the SP and IdP. It can provide the following benefits to the SP and IdP.

a) SP: The PIT requires registrations from the SP when it requests the PIT service the first time. The PIT administrator reviews the SP to filter out illegal collectors. The objects of this action are to reduce the number of illegal collectors and gain the trust of each SP to increase the self-disclosure level of the IdPs. The PIT provides the list of IdPs and their PI in an anonymity format that the SP can select as a target group in order to send the offer. This method can decrease the cost of the advertising because the SP can select quality IdPs as a target group and aims at increasing the response rate.

b) IdP: The PIT allows IdPs to approve or reject the incentive offers when receiving them from SPs. For this reason, it is possible to build up the IdPs confidence for providing real PI in the platform. The PIT decreases the amount of over-requested information by controlling the balance between an offered incentive and the requested PI. Moreover, the trading is not limited to the PI of the IdP but also includes the usage policy such as the usage time limit and information sharing agreement.

C. Trading Model development

Regarding the assumption that the PI trading activity constitutes a SP and an IdP, we propose a model that describes the trading of PI and associated incentives among SPs. We begin with the concept of a simple socially beneficial model to depict a comprehensive view of our model. We chose a socially beneficial model because money cannot be used to directly represent the value of PI. The IdP does not decide to trade his information solely for money. The social benefits for PI trading are the sum of the profits of both the SP and that of the IdP. Then, we also look at the concept of social exchange theory for looking into the decision of the trading activity for the IdP and SP because it suggests that a person's behavior is a result of the exchange process between the cost (C) and reward (R).

We have structured the simplest model for illustration purposes here. The situation is an SP who wants to trade an incentive to IdPs. Let i be the SP who wants to trade the incentive for the PI. The cost of the SP (C_{SP}) when providing an offer is the incentive value of the offer. The reward of the SP (R_{SP}) is the value from the PI including the value chain in the business. The total SP profit can be expressed by $\sum_{i=1}^{n}(-Ci_{SP} + Ri_{SP})$. There is only one SP in this study, so the total SP profit is

$$X = (-C_{SP} + R_{SP}). \quad (1)$$

Let j be the number of IdPs who use service provider i. The cost of the IdP (C_{IdP}) is the value of the PI from the IdP's perspective. The reward of the IdP (R_{IdP}) is the value of the incentives that were offered by the SP. When a SP sends an

offer of incentives to more than one IdP, the total profit of the IdP consists of the cost (C_{IdP}) and reward (R_{IdP}). It can be expressed as

$$Y = \sum_{j=1}^{m}(-Cj_{IdP} + Rj_{IdP}). \quad (2)$$

For the PIT, the value of the PI from the SP's and the IdP's perspectives are important. However, the values of PI from the IdPs' perspective always fluctuate and change. In the prototype design, each IdP has to complete a value of PI questionnaire before starting the trade. The results of the total profit of the SP (X) and IdP (Y) can be positive (+) and negative (-). Figure 3 shows a flowchart of a decision model in the PIT. When the SP can gain profit from the trade (X > 0), the PIT suggests it to create an offer. Then, after the IdP registers in the system, the PIT calculates the total profit of the offer for the IdP. If the IdP can gain profit from the trade (Y > 0), the PIT suggests the trade to the IdP. This decision module aims to create a win-win situation for both sides.

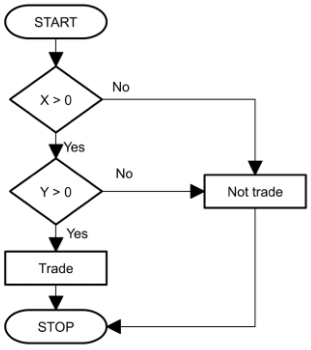

Fig 3. Decision flowchart of PIT

For example, we analyzed a situation in which the SP is in the hotel business. The hotel wants to increase the number of customers, so it creates a discount coupon to discount the room price to trade with the PI. C_{SP} is value of the discount + advertising cost. R_{SP} is value of the PI from the SP's perspective + the other related profits, such as the restaurant, spa, etc. C_{IdP} is value of the PI from the IdP's perspective. R_{IdP} is value of the discount coupon

D. Balanced personal information tradeoff Model

We propose a model for use in the PIT for estimating the balanced PI tradeoff in order to provide an effective and fair trade by supplying the maximum benefits to the SP while protecting the privacy of the IdP. This model comes from an assumption that each business has a different PI necessity level and each person has his or her own PI value. Each SP has a specified demand of PI attributes. The SP continuously increases the level of services or products when the IdP discloses their PI. However, the SP in each specified business has a limit of their services or products. When the IdP continuously provides more PI to the SP, the PI attributes will exceed the necessity of the business. In the trading, we focus on the value of each type of PI in a specified business (V_{SP}) and the value of each type of PI for the IdP (V_{IdP}). When V_{SP} is high, the necessity of that PI is high. When V_{SP} is low, the necessity for that PI is also low. On the other hand, the IdP does not want to disclose their PI when V_{IdP} is high. When V_{IdP} is low, the IdP easily discloses their PI. In this research, both the V_{SP} and V_{IdP} have been used to calculate the total value of the PI for a specified business.

$$Total\ value\ of\ PI = \sum_{i=1}^{n}(Vi_{SP} \times Vi_{IdP}), \quad (3)$$

where n is the amount of PI in a business.

Even if the IdP continuously provides more PI after all the necessary PI have been collected, the services or products of the SP remain the same. Thus, it is not necessary to provide PI that does not relate to the business. For this reason, the maximum PI tradeoff between the offer and value of the PI can be decided on from the highest necessity PI attributes for a specified business and the highest offer (100% of the offer). If the IdP wants to provide only parts of the PI, the offer can be decreased.

Moreover, there are several situations in which the IdP trades between the PI and the incentives. Both sides can possibly gain and lose their benefits. First, a SP may request a lot of information so the IdP loses their benefits when they trade high value PI for low value offers. Second, a SP may provide a high offer for low value PI. The balanced PI tradeoff model to avoid an unbalanced trade can be expressed using the following linear regression:

$$Offer\ level = \frac{1}{X_{max}} Total\ value\ of\ PI. \quad (4)$$

where X_{max} is a total value of PI at maximum offer (100%).

Figure 4 shows the relation between the offer level and the total value of the PI for a specified business on the PIT platform using the balanced PI tradeoff model. For the same amount of PI, a SP or an IdP might possibly lose benefit when they do not carefully make the trade.

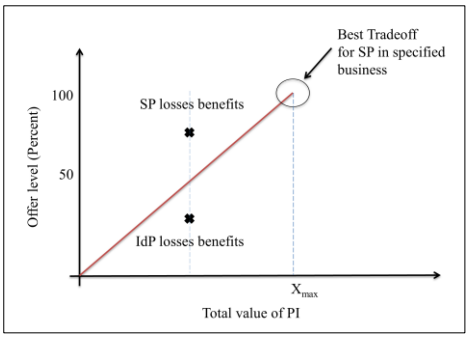

Fig 4. Illustration of balanced PI tradeoff

Moreover, we might be able to judge the value of the PI for creating a win-win situation by determining the maximum profit for both the SP and IdP. When a SP requests more PI from the IdP, there are three possible cases when we compare the value of the PI for the SP in a specified business and from the IdP during the trade for each PI attribute. Case 1: The value of the PI for the SP in a specified business is high but the value of the PI for the IdP is low. Case 2: The value of the

PI for the SP in a specified business is the same value as that for the IdP. Case 3: The value of the PI for the SP in a specified business is low but the value of the PI for the IdP is high. The PI in Case 1 is the best choice for the SP for gaining maximum profit during the trade because the SP can negotiate for a lower cost from the IdP. Thus, Case 2 is the second choice, and Case 3 is the worst choice.

V. A CASE STUDY: TRAVEL BUSINESS

In Section 4, we discussed the design of the PIT platform, trading model, and evaluation model. We aimed to evaluate our balanced PI tradeoff model in this case study. We have selected the travel business which is it is one of the most popular businesses that adopt online information technology into the business model as a case study.

A. Sampling frame and Measurement

The value of the PI for a SP in the travel business in this study was calculated by the necessity of the PI for the SP in the travel business. The following two sampling frame groups were the ones we focused on.

1. SP: We considered the necessity of PI in the travel business by focusing on the average value of PI in the travel business. We gathered the PI attributes from the registration pages of the websites from many businesses to study the necessity of PI for general SPs. The websites were selected from the top-ranking websites from the Alexa website, which publicly provides the website's ranking and statistics. We chose this data collection method because we wanted to find the minimum PI necessary for a specified business. We can interview SPs but they may add unnecessary PI in their answers for other objectives.

2. IdP: We used an online questionnaire to find the value of PI for the IdPs. In this research, we chose the comfort level in which discloser of their PI to a SP is acceptable as the value of the PI for the IdP. The sampling frames study came from people who regularly use the Internet. The monitors were in Thailand and they were between 15 to 70 year old. We used the five stages assessment system to ask about the comfortable level when disclosing their PI requested from an online SP. Level 5 is the most uncomfortable level for disclosing PI, whereas Level 1 is the most comfortable.

B. Results and analyses

Personal information was collected from 212 websites. There are a total of 35 PI attributes that were collected from the 212 websites for several businesses. From the websites collected, there were 25 travel websites and 19 PI attributes that were collected by the travel businesses. The number of websites that collected each PI attribute was calculated into a percentage. A higher percentage of the websites that requested PI means that the PI was more necessary in the travel business field. Then, the percentage was normalized from 0 to 5 for analysis purposes.

Next, we designed a set of questions for the IdP using the PI attributes that were collected by the SP. We asked for the comfort level when the IdP discloses each PI attribute. 532 people completed the survey. The respondent information in this survey is summarized in Table 1. The results for each type of PI were collected and calculated using arithmetic means for our analysis purposes.

TABLE I. RESPONDENT PROFILE

Age		**Education**	
15-20	3.7%	High school	5.4%
21-30	39.3%	College	2.8%
31-40	47.8%	Bachelor's degree	47.6%
41-60	8.6%	Graduate degree	40.5%
More than 60	0.03%		
Career		**Gender**	
Student	16.6%	Male	48.4%
Self-employ	14.0%	Female	52.1%
Private company	32.0%		
Government officer	27.2%		
State Enterprise	2.8%		
Unemployed/Housemaid	3.5%		
Others	3.5%		

The charts in Figure 5 show the results of the values of each PI attribute for the SPs and IdPs. The travel business collects 19 PI attributes that are maximum necessity PI for the travel business. The results were sorted and rearranged by V_{SP} from high to low. We can easily understand from the results the proportion of V_{SP} and V_{IdP}. In this case study, the values of some PI attributes are on the same direction. For example, the values for their home number and last name are important for both the SP and IdP. However, the values of some PI are totally conflicting. For example, the business phone number and company address have high values for the IdP but low ones for the SP.

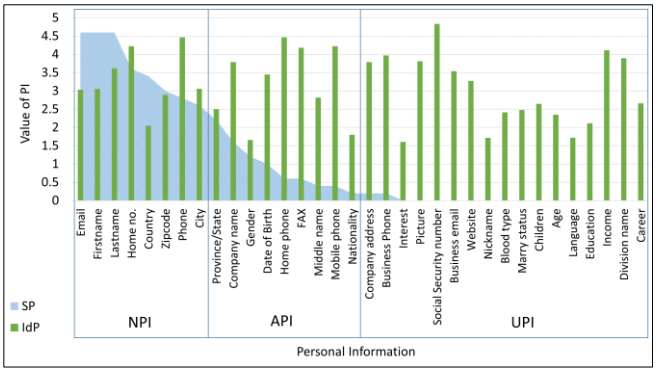

Fig 1. Value of PI for SP and IdP

Moreover, we are able to divide the results of PI in this case study into three groups using value of the PI for the SP. We assume a necessary point that separates the results when the V_{SP} is more than 2.

1) Necessary PI (NPI): The value of the PI for the SP is higher than the necessary point. These PI are highly required for a specified business.

2) Additional PI (API): The value of the PI for the SP is lower than the necessary point but higher than zero. These PI are moderately required for a specified business.

3) Unnecessary PI (UPI): The value of the PI for the SP is zero. These PI are of no value to a specified business.

The SP can reduce the trading cost by decreasing the request for some PI, which are API and UPI. If the SP needs more API, it can select API that has a high value of PI for them but a low value for the IdP, such as their gender.

Furthermore, Figure 6 shows the evaluation results from the balanced PI tradeoff model. The chart shows that the maximum tradeoff position when the SP collects all the necessary PI attributes and provides 100% of their offer. In order to reduce the offer cost, the SP may decrease the amount of PI that is requested from the IdP. Besides, SP can estimate the offer cost, if IdP negotiates to enclose some PI.

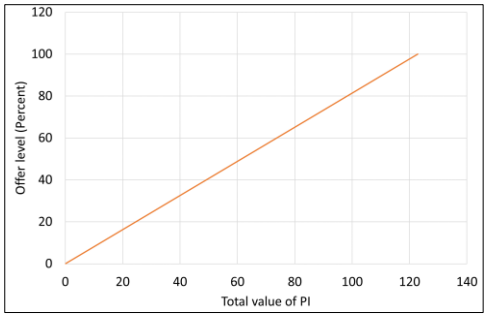

Fig 2. Evaluation results of balanced PI tradeoff model for the travel industry

VI. Conclusion and Future Work

Recent results suggest that PI collection is a major key to targeted advertising. Trading with an incentive has recently become a method that SPs use to attract the IdP to provide PI. However, the trade balance between the SP and IdPs is still difficult to adjust.

We tried to determine the current problems with PI collection and PI trading in this study. We proposed a personal information trading platform, called PIT, which aims at being a middle man between the SP and IdP. Not only does it provide the trading platform, but it also aims to recommend a reasonable trade providing maximum profit for the SP and privacy protection for the IdPs. We proposed a balanced personal information tradeoff model that uses the value of PI from both the SP's and IdP's perspectives to find the balanced tradeoff position. The balanced personal information tradeoff provides an appropriate level of trade in which the SPs do not lose their profits and the IdPs do not unnecessarily disclose PI.

The future challenges of the development of PIT are security issues and intelligence functions. Security technology has to be carefully integrated within the platform, because PIT contains PI that can be a target of criminals. After developing the platform, it will be possible to increase the system performance by combining it with an intelligence system such as an automated negotiation system.

VII. References

[1] R. T. Nimmer and P. A. Krauthaus, "Information as a commodity: new imperatives of commercial law," *Law and Contemporary Problems* 55.3, 1992, pp. 103-130.

[2] J. Frow, "Information as gift and commodity," *New Left Review*, 1996, pp. 89-108.

[3] C. Ciocchetti, "The Privacy Matrix," *Florida Journal of Technology Law and Policy*, vol. 12, no. 2, 2008.

[4] H. McCracken. (2013, June), *Cloudsweeper's Gmail Security Audit Is Alarming and Useful*, Available: http://techland.time.com/2013/06/27/gmail-security/

[5] M. Otsuki and N. Sonehara, "Estimating the Value of Personal Information with SNS, Utility," *Eighth International Conference on Availability, Reliability and Security (ARES),* 2013, pp.512-516.

[6] E. B. Andrade, V. Kaltcheva, and B. Weitz, "Self-disclosure on the Web: the impact of privacy policy, reward, and company reputation," *Advances in Consumer Research*, no. 29, 2002, pp. 350-353.

[7] E. Katrine, "Personal Information in New Zealand: Between a Rock and a Hard Place?" *Interpreting Privacy Principles: Chaos or Consistency? symposium*, Sydney, 2006, 1.

[8] Directive, E. U., "95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," *Official Journal of the EC 23* (1995): 6.

[9] S. 1332--109th Congress: Personal Data Privacy and Security Act of 2005, Retrieved March 4, 2014, from http://www.govtrack.us/congress/bills/109/s1332

[10] E. McCallister, T. Grance and K. Scarfone, "Guide to protecting the confidentiality of personally identifiable information". DIANE Publishing, 2010.

[11] O. Tene and J. Polonetsky, "Privacy in the Age of Big Data: A Time for Big Decisions," *Stanford Law Review* 64, 2012, pp. 63-69.

[12] A. Narayanan and V. Shmatikov, "Myths and fallacies of personally identifiable information," *Communications of the ACM*, 2010, 53.6: 24-26.

[13] A. R. Beresford, A. Rice, N. Skehin, and R. Sohan, "MockDroid: Trading Privacy for Application Functionality on Smartphones," *Proc. 12th International Workshop on Mobile Computing Systems and Applications, HotMobile '11*, 2011.

[14] G. Kontaxis, M. Polychronakis, and E. P. Markatos, "SudoWeb: minimizing information disclosure to third parties in single sign-on platforms," *Proc. 14th International Conference on Information Security*, 2011.

[15] J. Rosen, "The Right to Be Forgotten," *Stanford Law Review,* 64, 2012, 88-92.

[16] M. Otsuki and N. Sonehara, "A Proposal of "Identity Commons": Utilization of Life Log and ID Information for Resilient Social System," *2011 International Conference on 4th International Conference on Cyber, Physical and Social Computing*, *Internet of Things (iThings/CPSCom),* 2011, pp. 503-507.

[17] A. Felt. and D. Evans, "Privacy protection for social networking platforms," *Proc. Workshop on Web 2.0 Security and Privacy* (W2SP'08), 2008.

[18] E. Aïmeur and D. Schonfeld, "The ultimate invasion of privacy: Identity theft". *2011 Ninth Annual International Conference on Privacy, Security and Trust (PST),* 2011, pp. 24-31.

[19] K. Heather. (2012, August*), 83 million Facebook accounts are fakes and dupes*, Available: http://edition.cnn.com/2012/08/02/tech/social-media/facebook-fake-accounts/

[20] B. Wheeler. (2012, October), *Give social networks fake details, advises Whitehall web security official*, Available: http://www.bbc.co.uk/news/uk-politics-20082493