A STUDY ON PRIVACY TRADING: THE TRADING BETWEEN PERSONAL INFORMATION AND MONETARY INCENTIVES

Ake Osothongs

Doctor of Philosophy

Department of Informatics
School of Multidisciplinary Sciences
SOKENDAI (The Graduate University for Advanced Studies)

A STUDY ON PRIVACY TRADING: THE TRADING BETWEEN PERSONAL INFORMATION AND MONETARY INCENTIVES

Ake Osothongs

A dissertation submitted to the Department of Informatics
School of Multidisciplinary Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at
SOKENDAI (The Graduate University for Advanced Studies)

2017

A STUDY ON PRIVACY TRADING:
THE TRADING BETWEEN PERSONAL INFORMATION AND MONETARY INCENTIVES

Author:

Ake Osothongs

A dissertation submitted to the Department of Informatics
School of Multidisciplinary Sciences
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Informatics

School of Multidisciplinary Sciences

SOKENDAI (The Graduate University for Advanced Studies)

2017

A dissertation submitted to Department of Informatics,
School of Multidisciplinary Sciences,
SOKENDAI (The Graduate University for Advanced Studies),
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Advisory Committee:

Professor	Sonehara Noboru	National Institute of Informatics, SOKENDAI
Professor	Akihisa Kodate	Tsuda University
Associate Professor	Hitoshi Okada	National Institute of Informatics, SOKENDAI
Professor	Isao Echizen	National Institute of Informatics, SOKENDAI
Professor Emeritus	Shigeki Yamada	National Institute of Informatics, SOKENDAI
Professor	Yusheng Ji	National Institute of Informatics, SOKENDAI

ABSTRACT

This thesis focuses on a problem that how service providers exchange between consumers' personal information and monetary incentives, it aims to increase consumers' disclosure of personal information without increasing monetary incentives. The willingness to disclose personal information follows a complex process and each person values his or her personal attributes differently. Decisions as to whether or not to disclose personal information can be organized as a network structure. This thesis proposes a method to evaluate personal information using consumers' attitudes regarding personal attribute disclosure. The proposed method is used for our experiment by ordering the requested personal attributes. This thesis develops new knowledge to quantitatively increase the disclosure of personal information without increasing monetary incentives. Previous related works adopted many approaches such as auction and survey to assess the value of personal attributes; however, their results were only valid for specific situations. An adaptive approach is proposed here for more general situations. Although this case study selected Thai people as samples, by changing samples, this approach remains valid in a variety of situations such as in specific countries or within certain age ranges. Moreover, previous related works did not consider the dependency of personal attributes, whereas our thesis addresses the correlation of personal attributes from a more general approach, they can be considered as special cases under our approach.

The thesis consists of six chapters. Chapter 1 describes the background, problems, objectives, scope, limitations, preliminary definitions, and proposes study contribution. Chapter 2 outlines previous literature regarding the definition of personal information, privacy issues, problems concerning data collection, and assesses the valuation methods and notions of personal attribute monetary incentive trading. Chapter 3 compares the different viewpoints and trading angles between consumers and service providers. The collected data from Chapter 3 forms the datasets for Chapters 4 and 5.

Chapter 4 proposes a non-monetary valuation method for personal attributes. A graph is constructed based on Bayes' formula and analysed through graph mining techniques to determine relationships among personal attributes. Graph edges are used to compare values between each pair of personal attributes. Our graph proves robust within the evaluation context but encounters problems applying results in the absence of numerical values. Chapter 5 outlines the development of an application to conduct an experiment on the trading of personal information using monetary incentives. We propose a new technique to calculate the constructed graph into numerical values termed Value of Unwillingness to Disclose (VD). Personal attribute that contains high VD means consumers want to protect this personal attribute more than other personal attributes that contain lower VD. We then invite consumers who were separated into three groups to complete our evaluation. Each group is asked to decide their trade between personal attributes and prepares monetary incentives. The order of personal information by VD is arranged differently for each group as top-down from highest VD to lowest VD, bottom-up from lowest VD to highest VD, and adaptive ordered by consumer profiles.

Results indicate that it is possible to motivate consumers to disclose personal information without increasing monetary incentives. Participants disclose more personal data when the trading application requests personal attributes based on participant profiles. Chapter 6 summarizes the results and limitations and postulates directions for further research.

Our proposed approach can be used in different environments and for diverse groups of consumers; however, limitations and conditions are encountered during the study. To investigate personal information demands, we choose 212 top ranking global websites as our samples. Data regarding consumers' attitudes when disclosing personal attributes are collected from samples with similar perspectives toward personal information disclosure. Data are compiled from 532 Thai Internet users since the Internet and social media activity in the country rank amongst the highest in Asia. These datasets are incorporated into our proposed valuation method and 160 Thai participants are invited to complete the experiment. The proposed method of using personal attribute values to rank the order of personal information requests focuses on the negotiation mechanisms used on trading platform environments.

The knowledge created on the ordering of personal attributes can be used to improve the exchange of personal information through monetary incentive activities, such as requesting personal information in a survey and the creation of online questionnaires.

Currently, Thailand does not have any specific statutory law governing data protection or privacy; however, the government is in the process of drafting the Personal Data Protection Act. The findings from this study including personal attributes clustering, consumers' attitudes toward personal information, and ordering of personal information may be useful for organizations, and also relevant to the drafting of this Act in the areas of personal information categorization and personal data inquiry.

ACKNOWLEDGMENTS

The author would like to offer special thanks to his research supervisor, Prof. Sonehara Noboru, for his invaluable advice, guidance, and strong encouragement during the course of study. His support was invaluable to the completion of this study.

More especially, the author would like to express his gratefulness to senior supervisor and sub-supervisors Prof.Emer. Shigeki Yamada, Prof. Yusheng Ji, Prof. Isao Echizen, Assoc. Prof. Hitoshi Okada, and Prof. Akihisa Kodate for their useful recommendations, guidance, and practical advice. Furthermore, the author would like to convey his appreciation to his scholarship donor, the National Institute of Informatics (NII) and SOKENDAI (The Graduate University for Advanced Studies), for providing financial support during the study.

The author would like to express indispensable gratitude to the Department of Informatics, National Institute of Informatics supporters, staffs and faculty members for their kind assistance and support. The author would also like to thank his friends for all the support and encouragement offered during his study at NII. Lastly, the author would like to convey appreciation to his parents and family members, whose moral support and inspiration were instrumental in encouraging the author to pursue his studies tirelessly.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE TITLE **PAGE**

LIST OF TABLES

TABLE TITLE **PAGE**

CHAPTER 1

INTRODUCTION

Societies are now in the age of Big Data, where everyone has adopted smart devices into their daily life. The Big Data age is the information technology age, where tons of data are created every second from digital devices connected to the Internet. Many industries now rely on data from many digital sources. Service providers collect data from their consumers for many activities. Examples of activities where service providers rely on data are market analysis, target advertising, and product development. Along with many types of collected data, personal information is one of the important types of information that can be collected from consumers. Conversely the collection of personal information raises concerns of privacy problems in this Big Data age.

This chapter introduces our thesis. We describe the background, problems, objectives, and approaches of this study. We also provide contributions and a list of

publications related to the thesis. Lastly, we show the outline of the thesis, along with brief information on each chapter.

1.1 Background

Data are being generated continuously in the Big Data era. Data from many devices such as smart phones and personal computers contain personal information from consumers. Many types of service providers, such as research institutes, public organizations, and private companies collect and use personal information. These service providers base data collection on their need for personal information for several purposes, such as improving user experience, advertising, and research. However, the collection of personal information may lead to privacy intrusion problems. Consumers are increasingly concerned about their privacy because their personal information might have been collected without negotiation.

Service providers use many methods to encourage consumers to disclose their personal information. Many service providers use monetary incentives to persuade their consumers to disclose personal information. Examples of monetary incentives which service providers provide to consumers include discounts, coupons, and online services. Service providers may devote much of their marketing budget on monetary incentives to attract their consumers; nevertheless, consumers may not disclose their personal information because they consider the service provider's incentive insufficiently attractive. In general, consumers need high monetary incentive from service providers before disclosing personal information because consumers fear the invasion of their privacy. Conversely, service providers require as much personal information as possible, but do not want to provide high incentives to maintain control of their budget.

1.2 Problems

So far, the definition of *personal information* is an open question, and many debates have tried to define it. Definitions still vary because there are different opinions on what personal information consists of, such as technical aspects, culture, social rules, and local law. Therefore, many definitions of personal information exist with no standard definition, although many countries have different definitions of personal information for use in legal related activities.

Nowadays, service providers legally collect personal information through their systems when they receive agreements to collect and use personal information from each consumer. Service providers may collect consumers' personal attributes in their system, but it is complicated for consumers to manage personal information after it has been collected.

One of the challenges of this study is the method for estimating the value of personal information which can be used in an exchanging mechanism. The value of each personal attribute is difficult to estimate in currency terms, because consumers and service providers have different interpretations of the value of personal information. From the consumer's point of view, it is an asset, while from a service provider's point of view it is a resource. The willingness to disclose personal information follows a complex process. Each person values their personal attributes differently.

Service providers generally use monetary incentives to attract consumers to provide personal information. However, monetary incentives from service providers, and personal information from consumers are currently exchanged without an effective trading method. The question arises: how can service providers increase the disclosure of personal information from their consumers without increasing monetary incentives?

1.3 Thesis Objectives

The exchange between service provider incentives and consumer privacy is a major problem addressed in this thesis. The following tasks are addressed:

1. Comparison of service providers and consumers point of view toward personal information
2. Establishment of a method for estimating the value of personal information without considering monetary value
3. Development of an exchange mechanism which increases the disclosure of personal information from consumers without increasing monetary incentive

1.4 Scope and Limitations

There is no best solution for finding the optimal balance between monetary incentives and privacy disclosure when exchanging monetary incentives from a service provider and personal information from consumers. This thesis focuses on the perspective of service providers who initiate the exchange activity by creating an offer and offering it to consumers. The aim is not to increase monetary incentives provided to consumers, while increasing their personal attribute disclosure. The proposed method of using personal attribute values to rank the order of personal information requests is focused on the negotiation mechanisms used in trading platform environments.

Previous authors adopted many approaches to assess the value of personal attributes; however, their results are only valid for specific situations. This thesis aims to propose a general model that can be used in different environments and groups of consumers. However, limitations and conditions are encountered during the study. We have to limit the study to a specified group of consumers and service providers. We choose top ranking global websites from many businesses as our samples to investigate personal information demands. Consumers' attitudes data when disclosing personal attributes is collected from samples with similar perspectives toward personal information disclosure. Data is compiled from Thai Internet users since Internet and social media activity in Thailand ranks amongst the highest in Asia. These datasets are incorporated into our proposed valuation method for personal information. Moreover, Thai participants are invited to partake in the experiment. The sample group is drawn from Thai nationals only. Therefore, the results may possibly only be applicable to Thai nationals. Other groups with different cultures can evaluate personal information differently and the result can be different. Nevertheless, the methodology proposed in this thesis can be used to repeat the experiment for another cultural group in order to acquire an accurate result.

1.5 Preliminary Definitions

For the purpose of this study, the term *personal information* means any information relating to an individual. This can be any information, directly or indirectly collected from an individual, regardless of its source. The term *personal attribute* is also used when specified to any type of personal information.

1.6 Thesis Contributions

In previous researches, the value of personal information usually expressed in currency form. This thesis establishes a new method for valuing personal attributes and offers the possibility of showing relationships among them without considering currency. The calculated value of personal information shows that consumers value their personal attributes differently, and proves that it is possible to show the order of personal information disclosure in a hierarchy. The results of this work can be extended to other related studies. For example, many researches related to privacy disclosure has considered personal information as an equal value.

Consequently, service providers currently exchange monetary incentives with personal information from consumers without an effective trading method. This thesis develops new knowledge about ordering personal information requested, and shows how the value of personal information can quantitatively affect consumer personal information disclosure when exchanging monetary incentives for personal information.

Moreover, our proposed method of using personal attribute values to order the graph of personal information requests is specified to the negotiation mechanisms used on trading platform environments studied in this thesis. It is possible to extend the proposed method to other studies in different situations where personal information is required from consumers, for example, the requesting of personal information from online surveys.

1.7 Thesis Outline

The following is a description of the content of each chapter:

Chapter 2: This chapter describes works related to this study. We discuss the definition of personal information, the current situation of privacy issues in the big data age and problems relating to personal information collection. We also examine works relating to the valuation of personal information method. Lastly, we discuss personal information - monetary incentive trading.

Chapter 3: This chapter presents a comparison of the different points of view about personal information from consumers and service providers, namely, the demand of each personal attribute from service providers, and the importance of each personal attribute for consumers.

Chapter 4: This chapter presents our proposed method of personal information valuation. A graph is constructed and analysed to find the relationships among personal attributes.

Chapter 5: This chapter presents an improvement of the proposed method of personal information valuation from the previous chapter. Then, it describes the development of *Value of Unwillingness to Disclose (VD)* and how it is used for improving trading activities. Lastly, it presented the experiment.

Chapter 6: This chapter summarizes the results, limitations, and offers a future direction for this thesis.

1.8 List of Publications

Parts of this thesis have been published in the following publications:

1. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara, *Privacy Disclosure Adaptation for Trading between Personal attributes and incentives,* Journal of Information Processing, Vol.25 No.1 (Jan. 2017), page 2-11, 2017.
2. Ake Osothongs and Noboru Sonehara. *A Proposal of Personal Information Trading Platform (PIT): A Fair Trading between Personal Information and Incentives,*

International Conference on Digital Information and Communication Technology and its Applications (DICTAP 2014), page 269-274, 2014.

3. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara. *Evaluating the importance of personal information attributes using graph mining technique,* International Conference on Ubiquitous Information Management and Communication (IMCOM 2015), 8 pages, ACM, 2015.
4. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara. *A Prototype Decision Support System for Privacy-Service Trading,* The First IEEE International Conference on Multimedia Big Data (Big MM 2015), page 282-283, IEEE, 2015.
5. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara. *A Proposed Method for Personal Attributes Disclosure Valuation: A Study on Personal Attributes Disclosure in Thailand,* International Conference on Information Technology and Electrical Engineering (ICITEE 2015), page 408-413, 2015.

CHAPTER 2

RELATED STUDIES

In this chapter, we studied the related work of personal information collection as well as previous work related to attempts to resolve the privacy problems on personal information collection. Firstly, this chapter discusses the definition of personal information. Afterward, it defines the preliminary definition of personal information for this study. Secondly, it discusses potential problems when people disclose their personal information and that personal information is collected by service providers. Thirdly, previous studies of the valuation method for personal information are discussed. Lastly, the chapter discusses previous studies concerning trade between personal information and monetary incentives.

2.1 Definition of Personal Information

Privacy protection in this digital age usually focuses on the protection of personal information. However, the exact definition of the term "*personal information*" remains unclear and continues to be discussed [1, 2]. Time may change the meaning of this term, as the term is commonly found in legal documents. It has been updated parallel with the

development of information technology. There are other terms that have been used broadly in the same meaning such as "*personal data*", "*private data*" and "*private information*". People usually use these terms interchangeably in the same or similar context.

The traditional definition of personal information was different in the age when technology systems were still offline, such as in the 19th Century through the early 20th Century. The database of each system was separate. The definition of personal information usually entails information that can identify a specific individual.

In 1980, The Organization for Economic Cooperation and Development (OECD) published a guideline concerning the collection and management of personal information, OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data 1980. It was adopted by OECD member countries on 23 September 1980. The content gave the definition for personal data as

> "any information relating to an identified or identifiable individual (data subject) [3]."

One of the well-known privacy protection directives, Directive No. 95/46/EC of the European Parliament and of the Council dated 24 October 1995, concerned the protection of individuals about the processing of personal data. The free movement of such data defines personal information as

> "any information relating to an identified or identifiable natural person ('Data Subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity [4]."

An Australian law which relates to privacy is the Privacy Act 1988. It defines personal information as

> "information or an opinion, whether true or not, and whether recorded in a material form or not, about an identified individual, or an individual who is reasonably identifiable [5]."

Hong Kong's Personal Data (Privacy) Ordinance which came into force in December 1996 defines personal data as

> "relating directly or indirectly to a living individual, from which it is possible and practical to ascertain the identity of the individual from the said data, in a form in which access to or processing of the data is practicable [6]."

The Canadian Parliament published the Personal Information Protection and Electronic Documents Act (PIPEDA) which aims to protect consumer's personal information in 2000. It defines personal information as

> "information about an identifiable individual [7]."

In addition, many other countries also define the term in the same way, as information that can identify, whether directly or indirectly, an individual or particular person.

In some countries, the term "personally identifiable information (PII)" is used in the same meaning to describe information that can identify an individual [8, 9]. In a traditional system, information technology is mainly offline and each system is individual. A system containing personal information that can identify individuals through such means as phone numbers, home numbers, and social security number is easy to manage and control because it stores data in only one database or system. Privacy concerns are created due to risks when information technology connects many individual systems to work together as a network and then is connected to the Internet.

Furthermore, the development of technology changes the ways that people interact with information technology. Nowadays, people have adopted digital devices into their daily lifestyles, producing tons of information every second. In the past, some information was not defined as personal information because it was difficult to trace back to a person. However, information in this age may be produced by a person directly or indirectly. Researchers have proven that small pieces of personal attributes in this age of 'Big Data' make it possible to trace information back and identify individuals [10, 11, 12]. These small pieces of personal attributes cannot be judged by the same rule as with PII [13]. In 2012, the European Commission prepared a new draft for EU data and security laws,

currently known as the General Data Protection Regulation (GDPR). The definition of personal information has been adapted to this new Big Data age as

> "Personal data is any information relating to an individual, whether it relates to his or her private, professional or public life. It can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer's IP address [14]."

In other words, every personal attribute can be called personal information in this age because it can be combined with other personal attributes to identify an individual.

2.2 Data Privacy in the Age of Big Data

Traditionally, most information systems used a standalone database, including hospital information systems, accounting information systems, and university information systems. They did not share data with other systems across a network. Personal information was processed and stored in a single database, which was easy to manage and protect. Service providers manually requested personal information about their consumers from these traditional information systems. Eventually, the technology changed with the Internet age. Information systems are now connected to the Internet. Further, people connect themselves to the Internet via personal computers and smart devices such as smart phones and tablets.

In the early 21st Century, Big Data has become a well-known term to describe a large amount of data. There have been various definitions used to describe Big Data. A well-known definition was described by Gartner, Inc.

> "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. [15]"

Many industries are now finding benefits from big data. They see opportunities from big data, which is produced accurately by their consumers. This data can be analyzed using many methods and provide new knowledge about consumers, which can provide a

competitive advantage for service providers. However, it also raises new concerns about potential privacy issues. Data can be misused and create privacy violations.

Moreover, the number of people who use social network services (SNS) has increased dramatically. The number of social media users nowadays is more than 1.6 billion [16]. People are not just connecting together on social network sites, they also upload personal information such as pictures, video clips, locations, and their activities onto the social network site [17]. When consumers disclose personal information to social network sites, they also increase the chances for privacy issues such as identity theft and cyberstalking [18, 19].

Additionally, the adoption of smart devices has also had an impacted by greatly increased the quantity of data [20]. Additionally, the concept of Internet of Things (IoT) is now well-known and the number of devices connected to Internet are steadily increasing. Gartner, Inc. estimated that the number of IoT will reach 20.8 billion objects and IHS Markit estimated the number of IoT objects will reach 30.7 billion by 2020 [21]. Not only that these IoT objects can generated a large amount of data over the Internet, but it also contains more personal attributes [22].

2.3 Personal Information Disclosure and Personal Information Collection

A large industry for consumers' personal information is created by the high demand for personal information. Currently, the number of data brokers is estimated over 4000 data broker companies [23]. People give up their personal information when they connect to the Internet. Service providers collect personal information to understand their consumers and be able to provide personalized services or products to them. Nowadays, many business functions rely on data collected from consumers. For example, targeted advertising needs personal information to conduct a marketing pitch to a specified group of consumers. Personal information is a valuable resource for both public and private organizations in this digital age. It sometimes has been referred to as the new oil in this century [24] since raw data can be compared to crude oil, and we need to refine it to gain the hidden value.

To collect personal information traditionally, many websites provide online forms asking for personal information from their consumers. Moreover, service providers can collect data from public sources for some information. There is more demand for

consumers' personal information by many industries that come from big data analysis activities. Current technology makes the collection of personal information even easier. Service providers possibly crawl the data using the web crawler on the Internet. Service providers can collect automated information about consumers such as IP addresses, click streams and operation system information which has been automated generated by the system.

The collection activities of personal information have become a common issue confronted by everyone in the online environment. Many collection methods have been selected for collection of personal information. Service providers can collect personal information themselves and/or buy it from data brokers [25, 26]. When service providers need personal information from consumers using traditional methods, online service providers usually collect that personal information directly using online forms (registration) that require consumers to fill in their personal information. Additionally, some personal information is generated automatically on the service side such as IP addresses, operation system used and the time zone, which can easily be collected without consumer awareness.

Personal information collection activities have become a new privacy concern since service providers have begun collecting personal information. The more personal information service providers collect, the greater the risk of misuse. Even though every person has the right to disclose or withhold personal information and regulations exist in most countries, illegal collection activities are always happening on the Internet. Today, not only do businesses and researches collect personal information from consumers, other firms such as governments and hackers also collect consumers' information. Consumers have to risk their privacy with many illegal issues such as identity theft, cyberstalking and misuse activities when disclosing their personal information.

In a physical environment, people can easily refuse when someone comes to ask for their personal information. However, it is more complicated in an online environment. Service providers can collect as much personal information as they want. It is difficult for consumers to negotiate the disclosure of personal information. Service providers usually provide only two choices for consumers, accept or reject. Consumers who want to use a service or product are basically forced to accept. Making a judgment about disclosure is more complicated. The only option that consumers can use to ensure the collection, usage and sharing of their personal information will be protected by service providers is the privacy policy published by each service provider. However, many studies have found that

consumers have a tendency to not read the privacy policies [27]. One study gave several reasons for why privacy policies are ineffective. Firstly, they are often difficult to read and understand due to complicated verbiage. Secondly, consumers believe that their privacy is protected because the privacy policy exists. Thirdly, they don't actually read it because it takes a lot of time. Fourthly, once they have read the privacy policy, consumers don't have any choice. Lastly, it is not clear how users would protect themselves, as they do not see any harm in providing personal information to such websites [28].

The following are examples of other problems in personal information collection and trading:

A. *Illegal collectors*

Even though privacy laws and regulations that deal with the collection of personal information have been published in many countries and are widely debated, illegal collectors are still a problem. There are many untrustworthy personal information collectors online, such as unknown application providers who ask for consumers' personal information when they install applications and apps that carry malicious software [29, 30]. In some cases, service providers also collected personal information without users' consent [31, 32] and some personal information collection activities of service providers has become illegal in some countries [33, 34]. Moreover, personal information is sometimes illegally collected by government agencies [35, 36].

B. *Lack of Fair Trading*

Trading in personal information means buying, selling or bartering personal information [37]. People usually focus on the protection of privacy for consumers, but trading benefits for the service providers are usually ignored. Some researchers suggest that consumers should hide their PI to protect their privacy. Alastair et al. introduced MockDroid, a modified version of the Android operating system that provides a way to return valid but incorrect information to the service provider [38]. 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 [39]. These are examples of customer protections that do not return any values to the service provider.

C. *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 service providers [40]. Nowadays, many websites and applications state their privacy agreements and show the opt-out option. However, people sometimes cannot control the opt-out request, and some data brokers do not offer the opt-out option for their users [41, 42]. Only a small number of consumers know that data brokers offer a voluntary opt-out option [43]. It is difficult to track the data usage when it is already disclosed [44]. For example, consumers' emails are illegally collected by web crawlers and illegally sold online on the black market. Another problem occurs when personal information has been collected and there are no statements as to the limit of usage time for the personal information.

D. *Unbalance Trading*

Service providers always request as much personal information as possible. They can request more information than necessary. One problem is that it is difficult to find a balance between the protection of privacy and the utilization of information [45]. We are always faced with this kind of request for services, such as a request on a mobile phone application and social network login. Felt et al. found that popular Facebook applications tend to require too much personal information when a consumer requests the use of their services [46].

E. *Fake Information*

Service providers can collect automatically generated information. However, it is still necessary to collect consumer personal information directly. Consumers may submit fake personal information for several reasons, such as to protect their privacy and prevent marketing [47]. Criminals can use it for criminal activities such as identity fraud [48]. Some create fake profiles to hide themselves when they use online services, such as social networking services [49]. Moreover, some professional advisers suggest people use fake information when they do not trust the service provider [50]. This fake personal information can be a method to hide their

identity on the Internet. However, it could be argued that this leads to a new problem when service providers use personal information for legal purposes.

F. *Laws and Regulations*

Nowadays, the development of information technology has alerted consumers about protecting their personal information. Laws and regulations have become stricter in many countries [51], which is a reflection of new technology that is being organized to protect citizens. In general, service providers must receive consent from their consumers when collecting and using their personal information. Sometimes, service providers cannot use personal information, even if it has already been collected.

2.4 Valuation Method for Personal Information

Previous studies suggest that personal information should be treated as a kind of commodity. Personal information becomes a resource that can be used within a company or sold to others [25, 52]. Personal information has also been discussed as to whether it is a new currency or not [53, 54, 55]. As some believed it can be used as a currency [56, 57]. Consumers believe that their personal information is a type of asset that they can use for negotiating or trading with others. However, they also believe that service providers improperly gain benefits from their data and privacy. Businesses should make more of an effort to provide information and inform consumers about the risks and benefits of trading their data [58].

It is difficult to accept that personal information is being treated as an asset because it is difficult to estimate its value. Consumers always trade their privacy by disclosing personal information for online services such as email, search engines and entertainment. Data brokers sell personal information such as names, phone numbers and email addresses to third parties. Even though personal information can be treated as a commodity, the value of such personal information remains difficult to calculate.

The actual value of personal information is still difficult to estimate because people do not disclose their information just for tangible incentives; they also disclose their personal information for intangible incentives. From many studies, the value of personal

information is varied. The value of personal information can be very high in one study, while very low in another.

A study from the Financial Times estimated personal information worth for each person using pricing data from the industry in the US [59]. The results showed that personal information worth for the average person was less than one US dollar. Personal information from a single person increases when a person has a turning point in their life or change in their background. For example, they need to find something new and demand it in order to protect their story. Data brokers typically sell personal information such as the email and contact information of many people in a pack at a very low rate. Service providers do not have to buy it for each individual at a higher cost.

On the contrary, the cost of personal information from a consumer's point of view is higher. A study by Compassed Intelligences surveyed more than 1000 U.S., U.K. and Canadian citizens and asked them to assign a value to their personal information. The results showed that the overall value of their information on Social Network Services (SNS) was between $62.79 and $106.40 [60]. Both studies show the fact that service providers and consumers may have different visions concerning the value of personal information. From the consumers' point of view, their personal information has high value no matter who they are.

There are other researchers who worked on the value of personal information. Their results remain varied. For example, researchers developed a tool called "Cloudsweeper", which aims at identifying the value of an email account. The email account value is calculated from the service account values that are associated with each email [61].

Otsuki and Sonehara estimated the value of personal information using a SNS utility. The results showed an estimated value for personal information based on the cost of protection for that information [2].

A survey from Trend Micro asked consumers from all over the world to set a specific monetary value to each personal attribute. The result showed the average worth of personal information is $19.60. The results showed that the worth of each personal attribute is different by country. For example, the average value of health and medical record are $82.90 for US respondents and $35 for European respondents. Photo and video valued are $26.20 for US respondents and $4.70 for EU and Japanese respondents. They concluded that US citizens value their personal information higher than other counties [62].

The results comparison for each method is shown in Table 2.1. These researches are just a few examples of different opinions about methods for calculating the value of personal information [60, 63].

Even if the value of personal information is difficult to estimate, service providers still offer incentives as a reward to consumers in order to trade consumers' personal information. These rewards possibly affect self-disclosure decisions.

Researchers have found that the voluntary disclosure of personal information can be increased when the service provider offers monetary rewards [64, 65]. Service providers generally attract consumers to disclose their personal information by using monetary rewards such as money, which tends to increase the willingness to disclose personal information and decrease the risk of false information [64].

People currently disclose personal information without actual applicable value. Sometimes, they disclose personal information for a high value service, yet sometimes trade it for nothing. This is widely known as the privacy paradox problem. This fact shows how difficult it is to estimate the true value of personal information.

Table 2.1 Personal Information Valuation Method Comparison

Authors	Methods	Results
Steel, et al. (2012) Financial Times	Estimated Personal information worth based on the analysis of industry. pricing data in the US.	very low (less than $1 for every attribute)
McCracken (2013)	Estimated Personal information worth from the cost of service account values associated with the email.	Low / High (Depending on email)
Staiano, et al. (2014)	Participants create an auction from their data.	Low (€ 2 for each attribute)
Burney, et al. (2014)	Created a survey asking respondents to assign a value to their identity data.	High (From $62.79 to $106.40)
Trend Micro (2015)	Asked consumers to set a specific monetary value to each personal attribute.	High (Average worth of is $19.60.)

In recent years, a consequence of data mining applications and other exploration purposes is the desire to share our personal information encoded as tabular information. To reveal tabular information while still preserving the privacy of the consumer, several methods have been introduced. Those include *k*-anonymity [66], *l*-diversity [67], and *t*-closeness [68]. To keep data privacy, these schemes hide some specific personal information. There are many efforts to try and minimize hidden personal information [69, 70], In those researches, all personal attributes are equally considered, hiding important attributes such as a phone number is considered to be similar to hiding less important attributes such as gender.

2.5 Personal Information- Monetary Incentive Trading

Personal information can be considered a kind of commodity. It also can be used within a company or sold to third parties. However, it is usually difficult to accept being treated as an asset. Trading situations are possibly separated into categories by the type of service provider and situation. These can be categorized into public, private, commercial, and crisis situations as showed in Table 2.2. The need for consent from consumers is different in trading situations, as shown:

Table 2.2 Situations of Personal Information Collection

Situation	Requester	Require of Consent
Public	Government	Require or Not require
Private	Private company	Require
Crisis	Government	Not require

When government agencies require personal information from their citizens, some agencies do not require the owner's consent to disclose that personal information [71, 72, 73, 74]. Conversely, private companies are required to obtain the owner's consent prior to the release of personal information. Additionally, government agencies and private companies may not able to collect personal information directly from consumers or data creators. They may obtain personal information from a third-party potentially containing weaker privacy protection regulation [75].

In the case of crisis situations such as disasters and criminal related issues, consent is not required. For example, when the disclosure is necessary to identify the individual in disasters [76, 77]. However, the privacy of users should be preserved. For example, researchers proposed a method to access personal information on smartphone devices during a crisis, whilst preserving the user's privacy [78].

Nowadays, people trade personal information disclosure and monetary incentives on the Internet. However, the balance of the trade is usually ignored. Normal trading is based

on agreement between the service provider and consumer. The trade occurs when the service provider offers a monetary incentive to the consumer for trading personal information. The consumer feels comfortable disclosing personal information for those incentives. Figure 2.1 displays a common situation for trading personal information for incentives, comprised of a one-to-one relation between consumers and incentives. The incentives can be monetary, such as with money and coupons, or as a percentage discount.

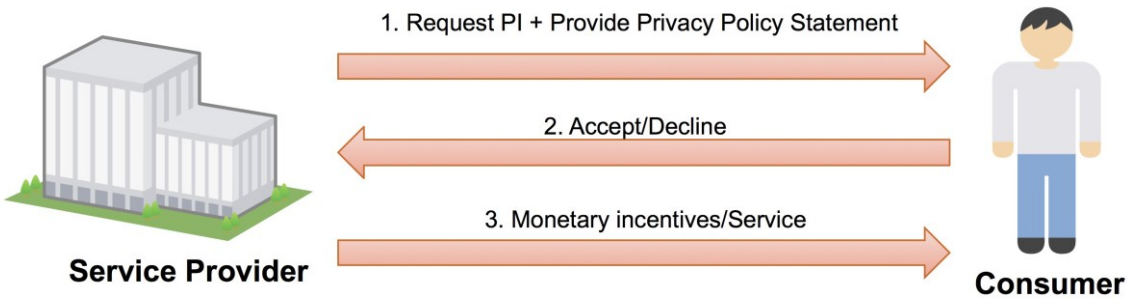


Figure 2.1 Current Personal Information – Incentives Trading Process

Presently, it is quite common for personal information to be traded for monetary incentives. However, the marketplace between service providers and consumers is rarely seen. Currently, there is some work related to the trading platform between personal information and incentives. Some startups provide services in the field of personal information trading with monetary incentives [79]. For example, Enliken, a company founded in 2011, provides an idea that allows consumers to exchange their data for discounts and donations [80]. Handshake focuses on a platform that allows consumers to exchange their personal information with currency [81].

Additionally, a trading platform discussed in previous work was proposed to support trading activities between personal information and monetary incentives [82]. The proposed platform was designed to contain three main components, including service provider, consumer, and personal information trading platforms (PIT). PIT was proposed as a platform to be placed between the service provider and consumer. Figure 2.2 shows the

platform architecture of PIT, which consists of three modules: security management, offer management, and incentive management systems.

The main perspective of service providers is to collect personal attributes, which are useful for their work. They will be able to provide monetary incentives in exchange for the personal information of consumers. We can see that these collections of activities, both online and offline, when service providers create campaigns which tradeoff between personal information and monetary incentives, such as discounts and online service. Service providers want to collect as much personal information as possible when exchanging monetary incentives for personal information. If consumers reject the offer, service providers will not provide anything to consumers. In other words, the assumption is that the monetary incentive is satisfaction for the service providers. On the other hand, the consumer perspective is more complicated. Even though consumers want to get monetary incentives for disclosing their personal information, they still want to disclose as little of their personal information as possible in return for high incentives. Consumers normally agree to provide unimportant personal attributes in trade for monetary incentives. However, consumers' concern for their privacy increases when service providers ask them to register or fill out their personal information directly, which reduces their overall satisfaction for providing their personal information. Consumers will often reject trading their personal information when they have very low satisfaction.

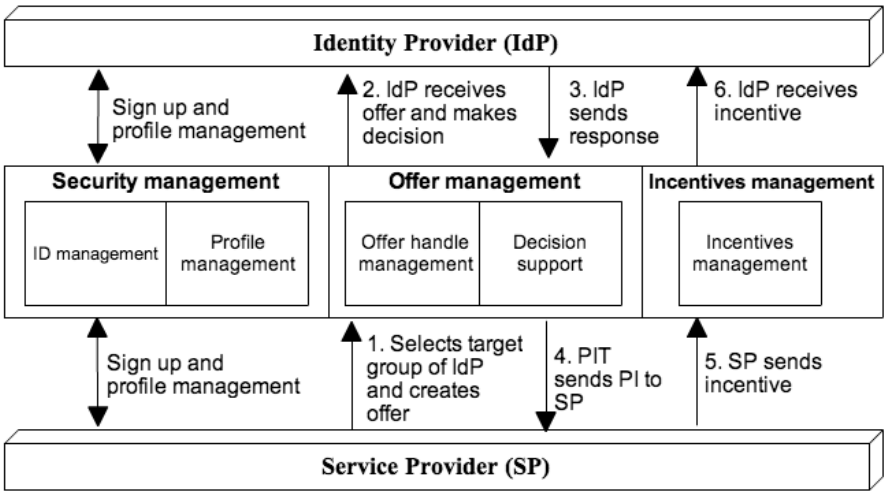


Figure 2.2 Platform Architecture of PIT

A problem with personal information trading is inside the trading method. Although personal information is possibly traded the same as other commodities and data markets for personal information already exist, there is still the lack of an effective trading method and negotiation mechanism. In 2002, the World Wide Web Consortium (W3C) officially proposed the Platform for Privacy Preference (P3P) [83]. It enables websites to reveal their privacy statements in a standard format, which can be interpreted by the web browser for delivery in a readable format for consumers. Each consumer can set up their privacy preferences when using a supported P3P browser. The browser will then automatically check the privacy statement of each website to avoid websites that do not match their privacy preferences. Even though the P3P standard is a well-known privacy protocol, its effectiveness for privacy protection has been questioned and critiqued [84]. This is because it lacks a negotiation mechanism. P3P was officially announced as a standard protocol many years ago, but there are few browsers in the market that support this standard. Therefore, previous studies proposed a privacy negotiation protocol [85, 86]. Many researches for negotiation of personal information have focused on protecting the privacy of consumer personal information during trade. This raises a new research question. When privacy protection is too high, the utility of personal information is low.

Additionally, Yassine and Shirmohammadi proposed a game theoretic negotiation method for the negotiation process [87]. They studied negotiation focusing on the trade-off between privacy risk and incentive in order to try and find Nash equilibrium. Ukil et al. proposed a framework that combined a negotiation-based architecture by using a prepared rule to create a negotiation matrix [88]. Moreover, Kwon proposed P4P (Pervasive Platform for Privacy Preference), a P3P extension using a multi-agent mechanism. It is a P3P-based negotiation mechanism for privacy management in pervasive computing services which allows users to negotiate in order to provide personal information following the user's privacy preferences [89].

CHAPTER 3

DEMAND AND DISCLOSURE OF PERSONAL INFORMATION

Personal information has become an important resource for most activities in the digital age. Service providers collect personal information from consumers and trade it with other online service providers. The trading activities between privacy and monetary incentives commonly have at least two important actors, the service providers and consumers. A service provider is an actor who creates the offer, while a consumer is the actor who receives an offer and must decide whether to accept or refuse it. The trading activity commonly starts from the demand of service providers who create an offer and then introduce it to consumers. Consumers then receive the offer and make a decision about whether or not to disclose their personal information. In order to improve the trading activities between personal information and incentives, it is important to understand the demand of service providers and disclosure the attitudes of consumers. From Chapter 2, the facts show that personal information is difficult to estimate and compare for its cost. The authors suggest that personal information value can be both tangible and intangible.

Therefore, cost estimation is not proper for personal value estimation. Service providers and consumers estimate the value of their personal information by using the importance of personal attributes from their point of view. Therefore, this chapter studies the different viewpoints concerning the value of personal information for service providers and consumers.

This chapter is separated into two sections, which includes study of the service providers' viewpoints and consumers' viewpoints. The first section is focused on service providers and aims to increase understanding of service providers' demand for personal attributes. This section studies the demand for each personal attribute from the top ranked websites. The second section focuses on consumers and aims to understand consumers' attitudes when considering whether to disclose personal attributes. It studies the comfort level of consumers when they disclose their personal attributes. Finally, the results from both sections are compared and discussed based on the study results.

3.1 Personal Attribute Demand from Service Providers

3.1.1 Overview

There are many types of personal attributes. Not all of them are related to or important to service providers. At present, service providers can request as much personal information as they want. Traditionally, personal information is collected using online forms such as user registration forms on websites and online order forms for e-commerce websites. Some of the personal information requested may not relate to the product or service offered by the service provider. Additionally, some service providers have adopted other methods to log in, such as social login services from social network service (SNS) platforms such as Facebook, Google, and Twitter. People who have an account with a SNS website possibly use their account to log into other websites with a few clicks. This login function has also become popular for mobile applications because users can log in to applications without new registration. For example, Facebook, one of the biggest SNS service websites, is used more than 850 million times per month. More than 60 percent of the top mobile

applications on Android used Facebook logins in 2013 [90]. However, consumers do not use the social login service without any cost; service providers commonly collect personal information from SNS service providers at the same time. Traditionally, consumers have to read and fill out a form while paying attention. Social network logins reduce the time for filling out and reading into a few clicks. However, the easier it is to provide personal information, the more risk of intruding on consumer privacy when consumers do not pay attention.

The first part of this study attempts to observe personal attribute valuation from the perspective of service providers. It also aims to categorize the personal attributes that service providers need to collect and the different types of service providers that collect them.

3.1.2 Methods

The authors of this study mainly focus on websites that collect personal information from consumers as service providers. However, the demand for personal information cannot be surveyed from the website owners directly because of the desire to avoid personal bias regarding demand for personal information. Website owners usually have ideas during the interview section and add some personal information that is not necessary for their business. Thus, this study started from selection of the sampling frame. A set of globally popular websites were selected as a sampling frame because these websites collect personal information from many consumers on a daily basis. The authors selected popular websites from the 'Alexa' website, which provides website rankings and visitor statistics for each website. However, underground websites were not studied, such as porn websites and illegal download software websites.

In this study, a total of the top 212 websites that were checked for the collection of personal attributes in 2013 were utilized and observed. There are many types of online service providers on these websites. The sampling websites are categorized by the type of online service provided. The types of services and their definitions are shown in Table 3.1.

Table 3.1 Types of services and descriptions for the sampled websites

Type of Service	Description
Bank/Finance related	The website focuses on banking services, credit cards, and financial information.
Business/Economy	The website focuses on business, economics, marketing, and business management.
Email	The website offers web-based email services.
Entertainment	The website provides entertainment services and information, including online entertainment services.
File storage/ Media sharing	The website provides online file storage and media sharing such as images sharing and file sharing.
News/Media	The website contains news reports and information, including for television and radio stations.
Newsgroups/Forums	The website mainly provides a bulletin board and newsgroup.
Reference	The website contains educational reference resources such as dictionaries and encyclopedias.
Restaurants/food	The website provides information related to food and restaurants.
Search engine/portals	The website provides search engine services and a web directory.
Shopping	The website provides goods or services. This includes the website itself, which provides advertising for various products.
Social networking	The website allows people to connect with others and share information with people in their network.
Sports/recreation/health	The website provides information about sports and recreation. This group also includes health-related information.
Technology/internet	The website provides information about technology and electronic devices.
Travel	The website contains information for travel planning, such as hotel reservations and booking, tickets and trip planning.
Weblog/hosting	The website provides a web community or web-based hosting services.

3.1.3 Results and Analysis

There are 47 personal attributes that service providers collected from customers on these websites. The results show that the demands for personal attributes from each website were different by the type of service. Each type of service provider had different requirements for personal attributes. Figure 3.1 shows the average number of personal attributes that were collected by the type of service. Bank and Financial related websites had the highest number of personal attributes. File storage/ Media sharing websites required the least number of personal attributes.

The results show that websites providing physical services and online services such as banking and financial service, shopping, and travel requested more personal attributes than the websites that provided online services only, such as news, file storage and references. The chart below shows the average number of personal attributes varies greatly.

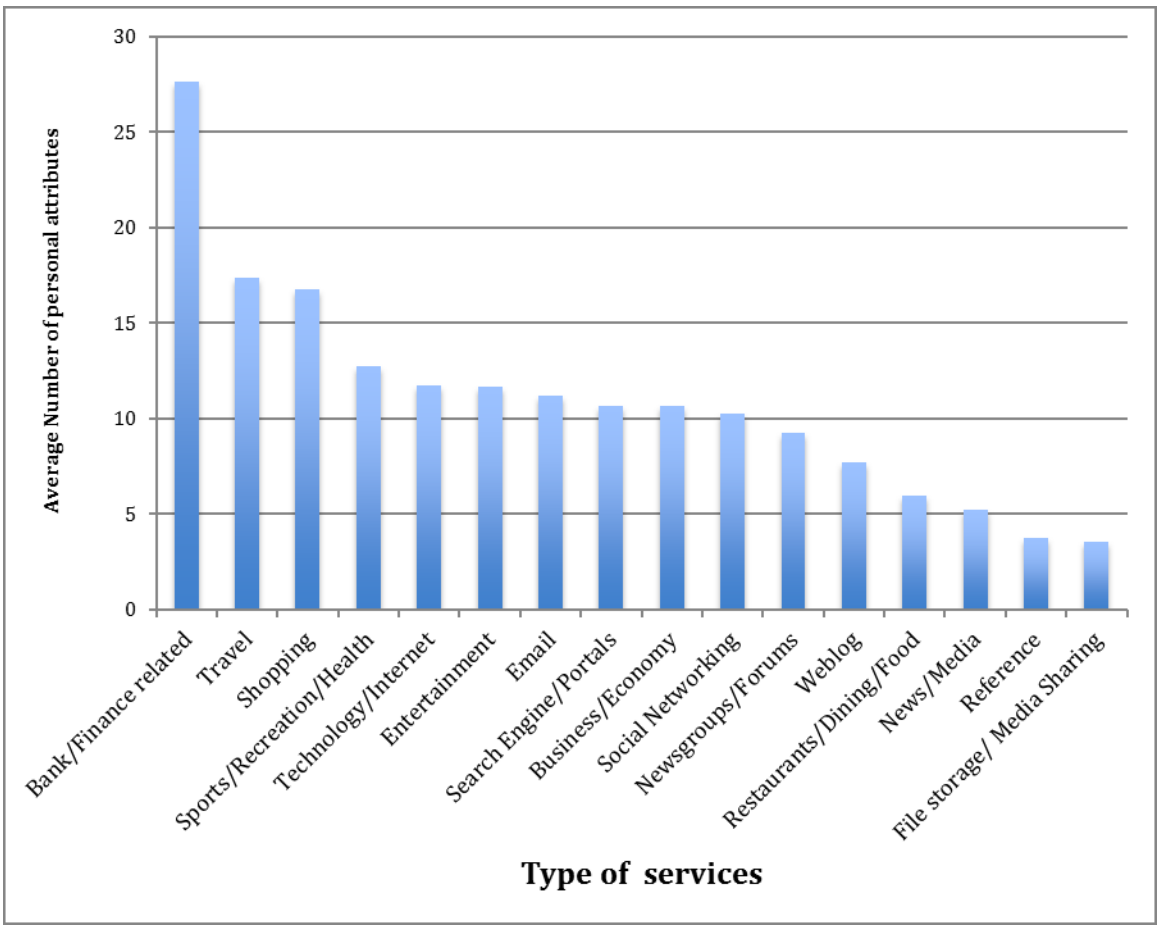


Figure 3.1 Demand for Personal Information by Business Type

This study analyzed the 47 personal attributes that were collected and explored the different types of personal attributes collected by the sampled websites in this study. For analytical purposes, the attributes are categorized by context into eight groups as follows:

1) Name information
   First name, Last name, Middle name, and Nickname
2) Demographical information
   Gender, Blood type, Marital status, Children, Nationality, Zodiac sign, Age, Language, Body type, Internet connection, Education, income, Location and Time zone
3) Home information
   Zip code, Country, Province/State, City, Building, Street and Home number
4) Company information
   Company name, Company address, Company type, Number of Employees and Division name
5) Phone information
   Home phone number, mobile phone number, Office number, Phone number specified to day time availability and phone number specified to evening availability
6) Email/Website information
   Email address, Mobile Email, Business email, Website, Email, Mobile email and Business email
7) Identifying information: Picture, Date of Birth, and Social Security Number or National ID number
8) Personal interests or preferences

Figure 3.2 shows the proportion of personal attributes that were requested by the type of service. Each service website collected different and obvious personal attributes. For example, the banking/ financial related websites collected the physical addressed of consumers, but email and reference websites did not need physical address information. Banking/ financial related websites did not require behavioral information about

consumers, but content websites such as sports/recreation, business/economy and technology/Internet website required information about consumers' interests.

To conclude, each specified type of service provider demands different personal information. The service-based service providers are flexible in their demands for personal attributes for both volume and type. In other words, the value of personal attributes from the point of view of service providers depends on their services. Moreover, the results chart also shows that the websites that provide physical services such as banking and travel information request identifying information more than websites providing online services only.

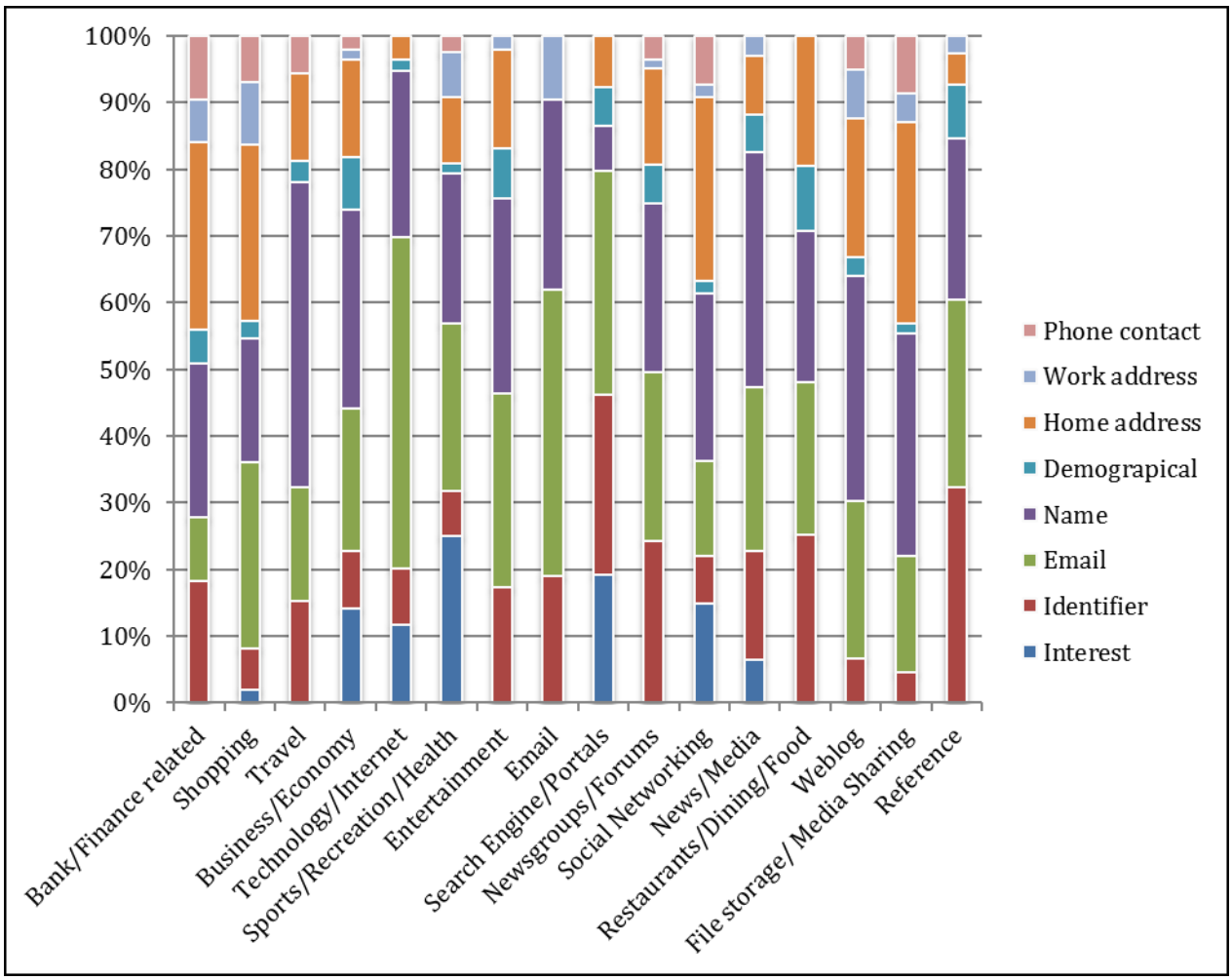


Figure 3.2 Percentages for collected personal attributes by service

Additionally, it was found that results which may relate to future work can focus on social login data collection. 69 websites were found to use the social network logins from 212 other websites. This means that more than 30 percent of the top-ranked websites are using social network logins. The number of social network logins is shown in Figure 3.3. Since the social login for Facebook has the highest adoption by website, the amount of information collected using traditional forms and Facebook login was measured.

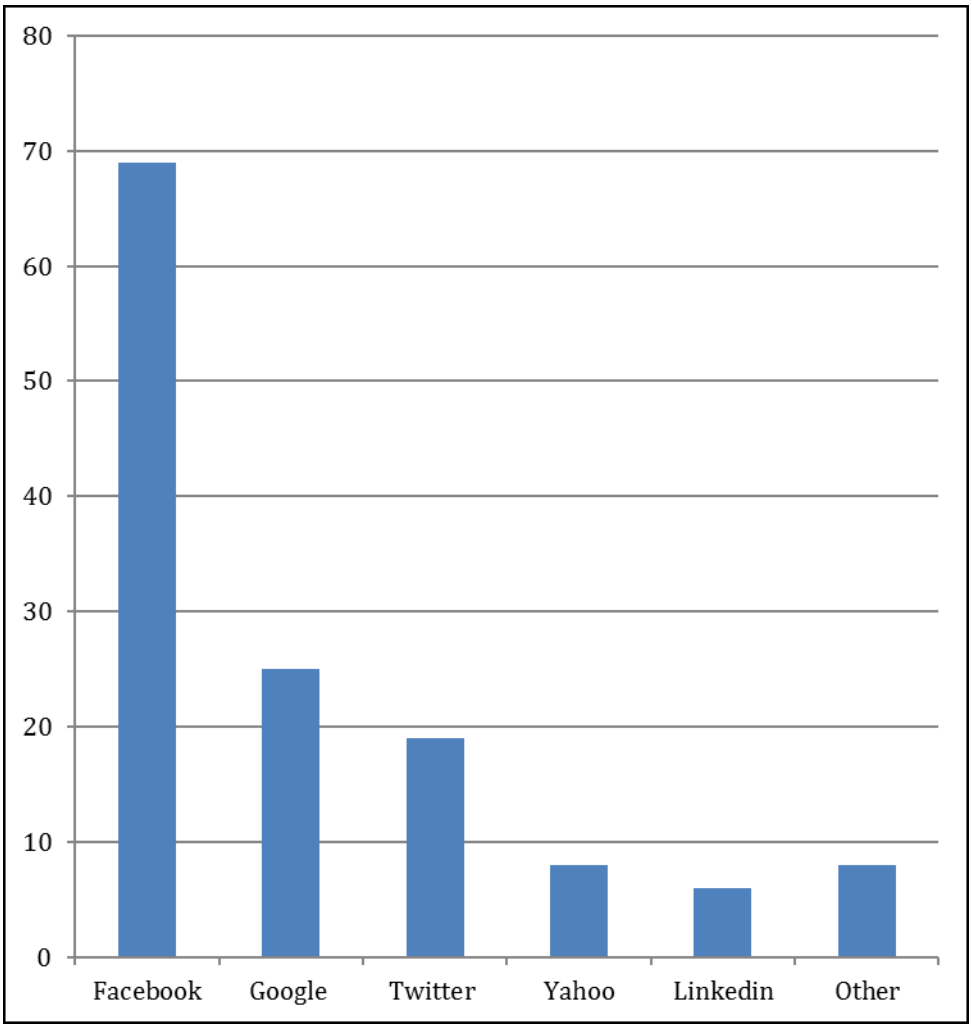


Figure 3.3 Amount of social network logins

Public information on Facebook is comprised of name, gender, profile picture, network, and cover photo. From this study, there are more than 30 attributes that service providers request from users. However, there is more information that social logins such as

from Facebook allow service providers to collect from users [91]. The authors compared the amount of personal attributes that service providers requested using social logins against online forms for the same websites. The focus was on the attributes involved in public profiles, emails, and birthdays. These are the top three results because there are many attributes that cannot be collected only using online forms. The results of comparisons are shown in Figure 3.4.

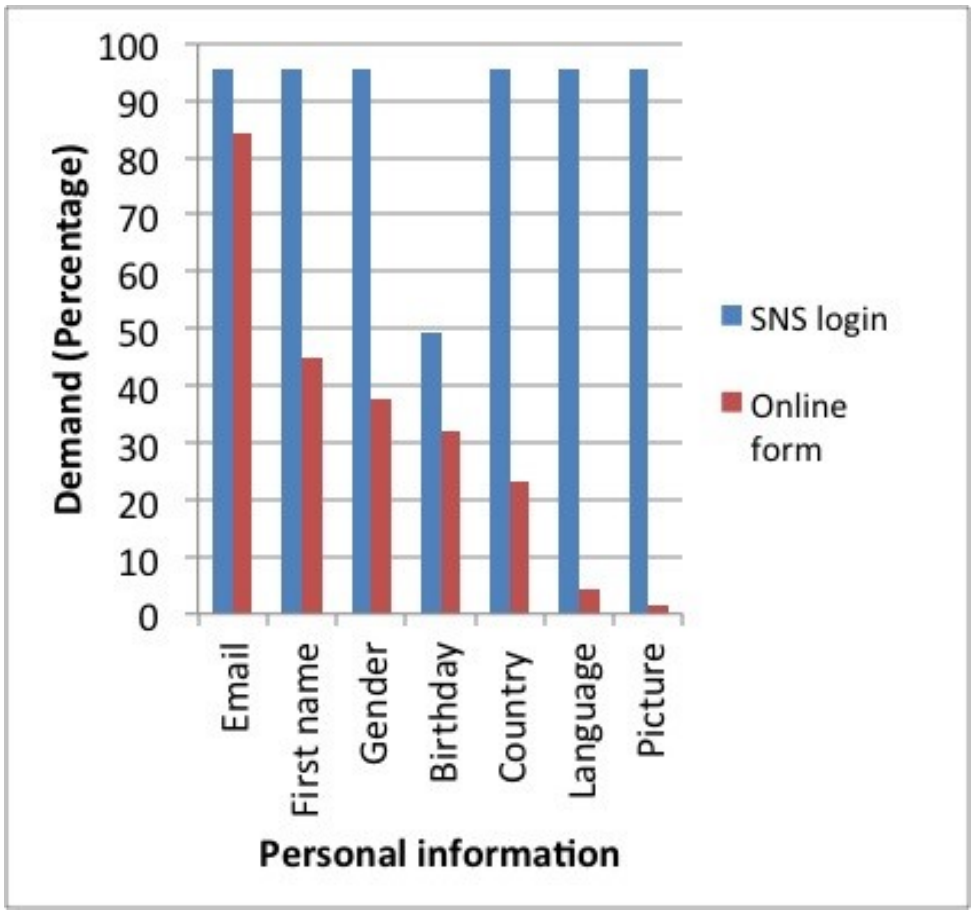


Figure 3.4 Comparison of requested personal information from SNS login versus traditional online form

The results of comparison show that, even on the same website, requested personal information from Facebook login was higher than from the online form login. The power of social networking provides information that cannot be collected when using the traditional form, such as their network list and check-in location. However, it also means service providers who can provide services to consumers using personal information collected

from the traditional online forms receive an excessive amount of personal information when consumers adopt social logins. They can disclose unnecessary personal information to the service provider without any intention to do so.

3.2 Personal Attribute Disclosure of Consumer

3.2.1 Overview

Section 3.1 showed the demand for personal information from service providers, which can reflect the estimation of personal information. This part continues the study of attitude for personal information in consumers' point of view. The related work in the previous chapter showed that personal information is difficult to estimate for value. The quantitative value of personal information is sensitive to its context. The value of personal information is not easily described by currency in consumers' point of view.

Thus, this study aims to find the relative value of personal attributes from users' point of view, which is possible using the trading activity between personal information and incentives.

3.2.2 Methods

This study used a set of questionnaires to study the comfort level of consumers when disclosing their personal information to service providers. A set of questions was designed to ask participants about their comfort level when disclosing each personal attribute.

The participants in this study were Internet users who usually disclosed their personal information online. The participants were in Thailand and aged between 15 and 70 years old. There were 532 participants that completed the questionnaires. The questionnaires had two parts comprised of demographic information and comfort level when the consumer disclosed each personal attribute. The participants' information in this survey is summarized and shown in Table 3.2.

Table 3.2 Information about participants

Personal Attributes	*Value*	*Percentages (%)*
Education	High School	5.4
	College	2.8
	Bachelor's degree	49.5
	Graduate school	42.2
Gender	Male	48.4
	Female	51.6
Age	15-20	3.7
	21-30	39.3
	31-40	47.8
	40-60	8.6
	More than 60	0.03

The second part asked about comfort level when disclosing each personal attribute using a 5-point Likert scale questionnaire. The highest level was 5, meaning the participant strongly disagreed with disclosing the personal attribute or the participant was uncomfortable with disclosing the personal attribute. The lowest level was 1, meaning the participant strongly agreed with disclosing the personal attribute or the participant felt comfortable in disclosing the personal attribute.

3.2.3 Results and Analysis

The questionnaire results were collected and calculated using arithmetic means for analysis purposes. The results are shown in Figure 3.5. The participant results showed their attitudes when disclosing personal attributes to service providers. The results also showed that the personal attributes that had the highest means were those that easily identified an individual, such as National ID number, home phone and mobile phone. Next, the personal attributes for contact information, such as address and email, had lower value. Finally, demographic information had the lowest value from the results.

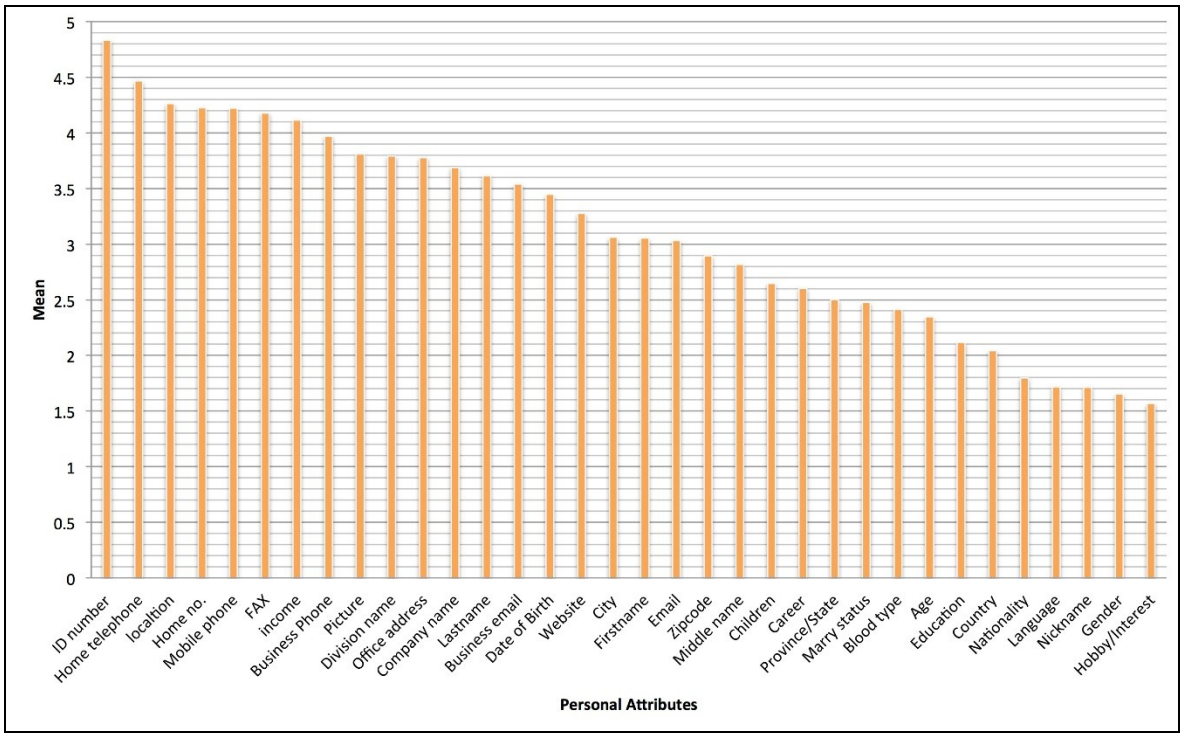


Figure 3.5 The survey results for each personal attribute

3.3 Comparison of Different Viewpoints for Personal Information

To clearly display the different viewpoints on personal information for service providers and consumers, the authors selected the travel industry as a case study for comparison purposes. From the study, travel websites required consumers to input a high amount of personal attributes, which they may use for reservations and related activities. From the collected websites, there were 25 websites from the travel industry. These websites required 19 personal attributes highly necessary for the travel industry. The study checked the number of websites which required each personal attribute from these travel industry websites and then compared their demand for personal attributes and the comfort level among consumers when asked to disclose personal attributes. For analytical purposes, the

authors normalized the numbers on a scale from zero to five. The compared results were sorted and rearranged, as shown in Figure 3.6.

From the results, the comparison showed that service providers' and consumers' demand on some personal attributes are in the same direction, while some are different. For example, some personal attributes such as home number and last name are important for both the service providers and consumers. However, the results of some personal attributes are totally different. For example, office number and company address have high value for the consumers, but are not required by service providers. It can be assumed that consumers felt that their office address was important for them, but service providers in the travel industry do not require the office address of consumers for their activities. Moreover, most demographic information is not required on travel industry websites.

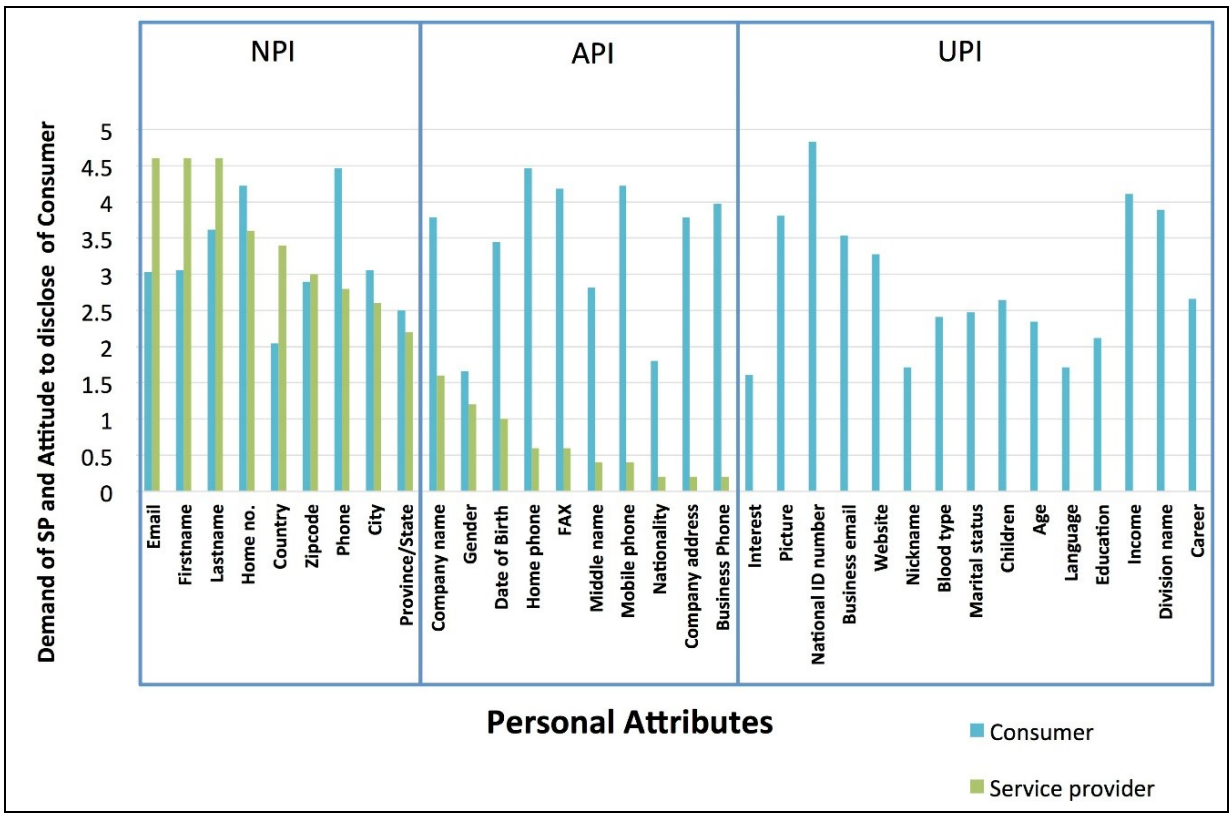


Figure 3.6 Demand for personal attributes from travel websites and attitude of consumers to disclose personal attributes

To sum up, the results showed different viewpoints for personal attribute valuation from service providers and consumers. The valuation from service providers was significantly larger compared to than that from consumers. Service providers had personal attributes with significantly higher valuation than others. On the other hand, consumers gave more similar valuation for most personal attributes.

3.4 Summary

The valuation of personal information is difficult to estimate in terms of currency since it is sensitive to context. Both service providers and consumers have different opinions about the valuation of personal attributes. The authors adopted these ideas and separated this chapter into three sections. The first section of this chapter studies the demand for personal information by service providers. Then, the second section studies the importance of personal information for consumers. The last section shows the different viewpoints on personal attribute valuation for service providers and consumers. The authors confirmed the assumption that service providers and consumers estimate their personal attributes differently. The valuation of personal attributes from service providers has significantly larger variation than those from consumers. On the contrary, consumers give more similar valuation for most personal attributes.

Moreover, the results of this chapter showed that even though service providers require many personal attributes for their services, they do not require all personal attributes from consumers. Each specified type of service provider has its own demands for personal attributes. From the results of comparison in Figure 3.6 for this study, the authors separated personal attributes from the case study into three groups using service providers' demand for personal attributes. It is assumed that a necessary point separates the results when there are more than 2. The three groups are:

1. Necessary personal attribute (NPI): an important personal attribute from service provider's point of view. This type of personal attribute is highly required for a specified business.

2. Additional personal attribute (API): a personal attribute which is sometimes important from service providers' point of view. This type of personal attribute is moderately required for a specified business.
3. Unnecessary personal attribute (UPI): a personal attribute that is not important from service providers' point of view. This type of personal attribute is not required for a specified business.

The understanding on viewpoint for personal information can be used to improve the trade-off between personal information and incentives. Since the results showed that consumers commonly estimate their personal information value higher than service providers, service providers may reduce incentives by decreasing the amount of requests for personal attributes belonging to the API and UPI groups.

CHAPTER 4

A PROPOSED METHOD FOR PERSONAL INFORMATION VALUATION

4.1 Proposed Method for Personal Information Valuation

At present, both service providers and consumers are more concerned about the value of personal information than in the past. Service providers look at personal information as a type of asset and demand it as much as possible. Meanwhile, consumers do not want to disclose it without benefit [58]. However, it is difficult to trade personal information without understanding each personal attribute's value. This chapter proposes a valuation method in order to evaluate the value of specific personal attributes and uses it to develop an application as a case study.

4.1.1 Overview

There are many researches that assume all personal attributes have the same value and should be considered equally. On the other hand, various work has tried to estimate the worth of personal information using different methods. The results vary as discussed in Chapter 2. Each method has different results, which reflect that the value of personal attributes is sensitive to context. Personal information can be determined as both tangible and intangible. Therefore, service providers and consumers estimate the value of personal information from different points of view. The results of the questionnaires in Chapter 3 showed that the value of personal attributes varied according to consumers' point of view. For example, the results showed most consumers estimated the value of personal attributes such as National ID as having more value than age or gender.

In order to complete the negotiation process, service providers need to understand how consumers value each of their personal attributes. Therefore, the authors proposed a personal information valuation method to gain understanding of each personal attribute, which could possibly be used in the trading activity in this chapter. The authors found that many studies proposed cost estimation methods in order to identify the value of personal attributes. However, this study found that their results relied mostly on the value of each personal attribute and could not show relationships among personal attributes. In addition, the results of previous studies motivated this study to estimate personal information value by using its' context. The authors used consumers' attitudes to disclose each personal attribute in Chapter 3 to estimate personal information value. In the negotiation process of the trading platform, the selection of personal information cannot be based solely on individual value. When a consumer rejects disclosure of a personal attribute, the next personal attribute will be offered to consumers. The mechanism for selecting the next personal attribute must be designed. The relationship between personal attributes is necessary for selecting the next personal attribute in this case. Therefore, this chapter aimed to find the relationships among personal attributes from consumers' point of view and proposed a valuation method in order to evaluate the value of personal attributes used in this study.

This chapter proposes a technique to measure and compare the value of each personal attribute without quoting financial value. The proposed method used the questionnaire results to construct a graph based on attitudes when consumers disclosed their personal attributes from the previous chapter. Then, the authors used a graph-mining technique to extract relationships between personal attribute disclosures. The results showed an interesting hierarchical relationship between personal attributes. Further, it was found that consumers tended to protect personal attributes that have semantically similar meaning.

This section is organized as follows: Firstly, the research methods, calculation method and graph method are detailed and described. Secondly, the authors calculated a priority of personal information and then created a graph for personal information disclosure. Thirdly, the authors extracted the community of the graph and showed the results. Finally, the use of the proposed method in a prototype of a decision support system is shown.

Graphs shown in Figures 4.1, 4.2 and 4.6 are visualized using Gephi software [92].

4.1.2 Personal Information Valuation Visualizing

The results from previous researches showed that personal information value was sensitive depending on context. There have been many types of surveys used to estimate the value of personal information, as discussed in Chapter 2. Many researchers have tried to estimate personal information value using financial value. These studies possibly ranked the value of personal information from highest to lowest. However, it is difficult to use these studies' results for trading activities. Even if it was possible to compare personal attributes using their financial value, this study could not use those same values while ignoring context. For example, the worth of two personal attributes, email, and age, may have equal financial value in some studies, but different groups of consumers may judge their personal attribute value differently. Moreover, these works also could not display the relationship value among each personal attribute since the relation of value between personal attributes contains complicated information. In this study, the graph was chosen for displaying the

relationships of value between each personal attribute. A graph provides several advantages when comparing personal information in a trading platform because it allows displaying the relation of personal attributes in a hierarchy and easily compares multiple personal attributes.

4.1.3 Development of Proposed Valuation Method

In the previous chapter, the authors collected the demand for personal information from 212 popular websites. It was found that most of the personal attributes were generally collected from those websites, but some personal attributes are too specific to a certain website. In this chapter, the authors focused on 33 personal attributes commonly collected by service providers. The method used to collect data from the samples were Likert scale five-level questions, which asked for comfort level when consumers disclosed their personal information. The highest value was 5, meaning strongly disagreed with disclosing their personal attribute. In other words, consumers felt uncomfortable when they disclosed that particular personal attribute. The lowest value was 1, which could imply that consumers felt comfortable when they disclosed that particular personal attribute. This study separated the results into two conditions, which were *'disclose'* and *'protect'*. *'Disclose'* means the consumers disclose personal information to service providers. *'Protect'* means the consumers do not agree with disclosing personal information to service providers. When the answers to the questions were 1, 2 or 3, the condition was *'disclose'*. When the answer to the questions was 4 or 5, the condition was *'protect'*.

This study calculated the probability for protection of a personal attribute given the other personal attributes. This study adopted Bayes' formula in calculating condition probability.

Let b be the personal attribute given by personal attribute a, the calculation is expressed by:

$$P(b \mid a) = \frac{P(a \cap}{P(a)} \quad (1)$$

where $P(b \mid a)$ is the probability that consumers protect personal attribute b when consumer already protect personal attribute a. $P(a \cap$ is the probability that consumers protect both personal attribute a and b. $P(a)$ is the probability that the consumers protect personal attribute a. It calculates $P(a \cap$ by the number of subjects who chose to protect both a and b divided by the number of participants in the questionnaire, while $P(a)$ is the number of participants who chose to protect a divided by the number of all participants.

The dataset of calculation results contained pairs of relationships for disclosure among personal attributes. When $P(b \mid a)$ is close to 1, almost all of the participants who protect personal attribute a also protect personal attribute b. In other words, personal attribute a has more value than personal attribute b in the consumers' point of view.

After calculating the probability of protecting personal attributes, this study transformed the results dataset into a directed graph to visualize dependency among personal attributes. The study defined personal attribute a as more valuable than personal attribute b when consumers want to protect personal attribute a after protecting personal attribute b. Each personal attribute is represented by a node of the directed graph.

Each edge of the directed graph represents a relation between two personal attributes. Each pair of nodes is linked from a high valuable node to a low value node. Every edge also contains a weight calculated using the probability of personal information disclosure between two connected nodes. The direction between nodes shows the priority of the disclosure for each personal attribute. Each pair of personal attributes has a parent node and child node. Therefore, the parent node is the node that has higher value from consumers' point of view. The constructed graph is shown in Figure 4.1.

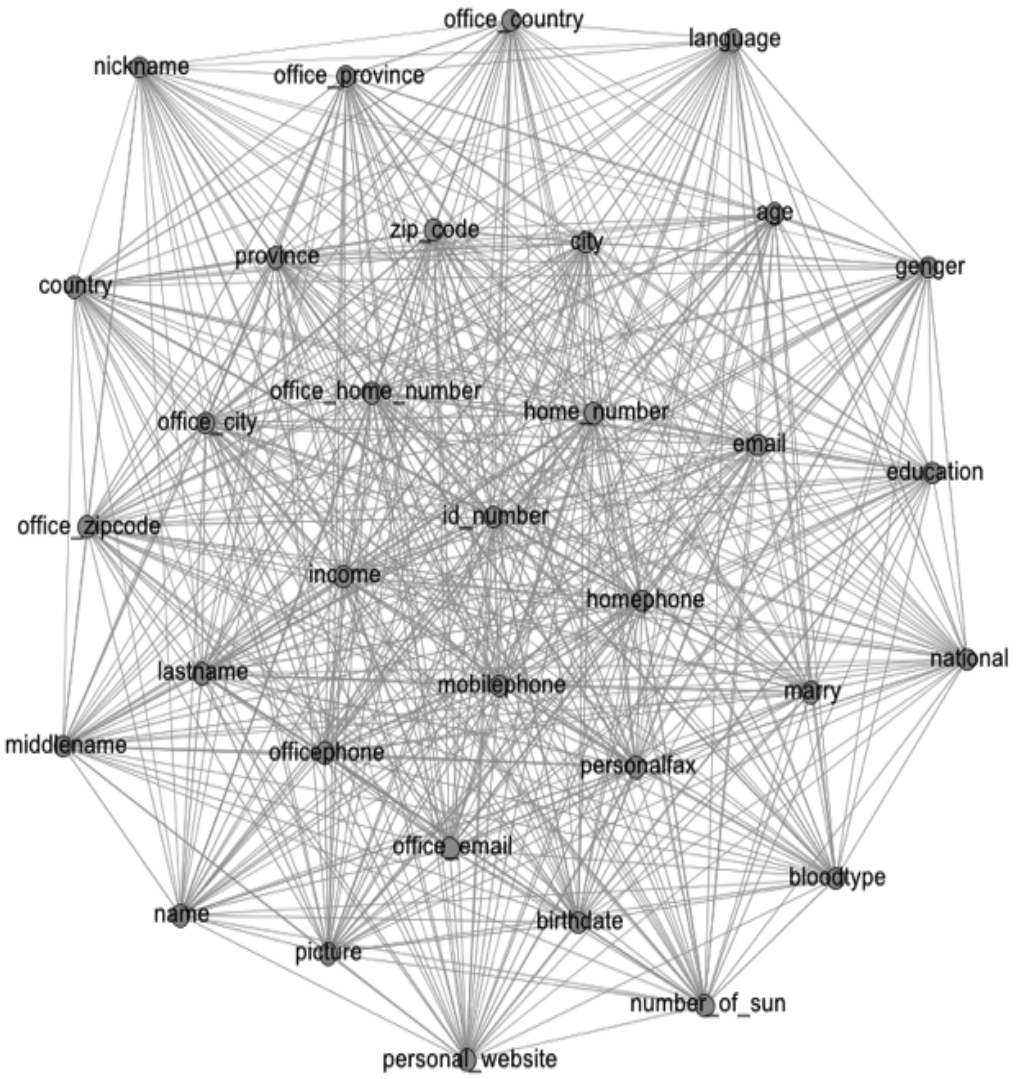


Figure 4.1 Initial direct graph containing relations between personal attribute disclosures

The direct graph in Figure 4.1 was still complicated to understand from current visualization and was difficult to use practically because it showed all relations, including which personal attribute relates to all other personal attributes. This study eliminated some edges from the graph based on the idea that almost all of the participants who protected personal attribute a also protected personal attribute b if $P(b \mid a)$ was close to 1.

Let α be a probability close to 1, the authors eliminate edges (a,b) when $(b \mid a) < \alpha$. The resulting graph after removal is shown in Figure 4.2. The results in Figure 4.2 show

only the personal attributes that have a tight relation between them. In this study, set $\alpha = 0.95$, and almost the same results were achieved when setting $0.9 \le \alpha \le 0.95$ in this experiment.

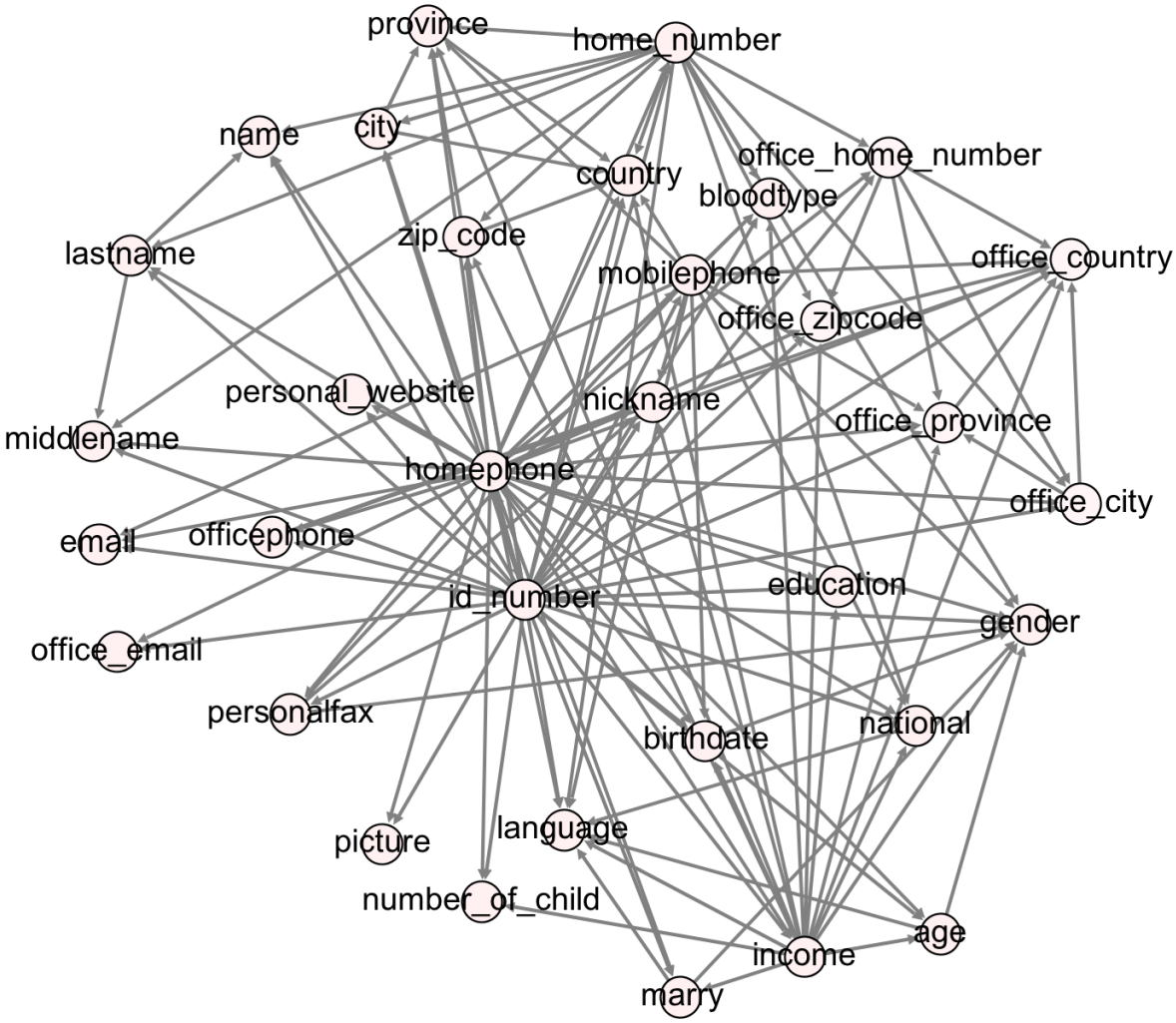


Figure 4.2 Direct Graph displaying the Relation between Personal Attributes

Considering the distribution of in-degree and out-degree for all personal attributes, the authors found that it contained meaning. The node that had more out-degree connected to other nodes could be perceived as an important personal attribute in consumers' point of view. The study found that some personal attributes such as *National ID number*, *home phone*, and *mobile phone* had more out-degree from their nodes connected to other nodes than most of the personal attributes. It could be implied that most consumers paid more attention to *National ID number*, *home phone* and *mobile phone* when compared to other

personal attributes. On the other hand, it was also found that some personal attributes such as *age*, *gender*, *language*, and *nickname* were considered less important from consumers' point of view because they had no out-degree at all.

In order to measure the value of each personal attribute, the authors considered the in-degree and out-degree of each personal attribute. The results showed that once people disclosed one of these personal attributes that had higher out-degree, they could disclose more personal attributes that had lower value than the first personal attribute in the hierarchy.

For example, the results for selected nodes are shown in Figure 4.3. There were five nodes in the graph, including *home phone*, *National ID number*, *home number*, *city,* and *province*. From the level of hierarchy, it could be implied that *home number* was one of the most important personal attribute in consumers' point of view. It also had the highest value in this node. Therefore, most consumers that disclosed their *home number* also had a high possibility of disclosing their *city* and *province*. Moreover, they also had a high possibility of disclosing their *city* and *province* when they disclosed *home phone* and *National ID number*, as both of these personal attributes were in the lower level.

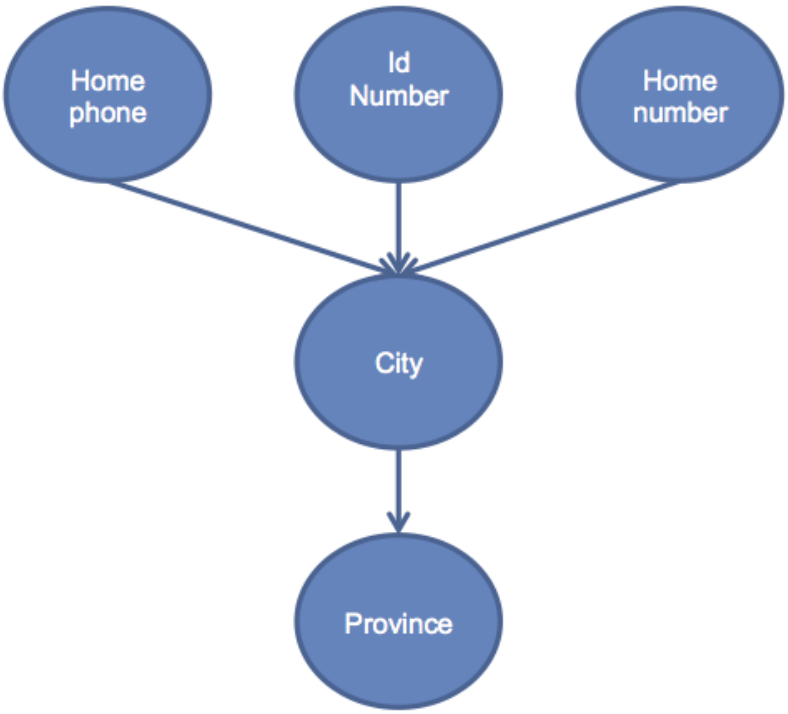


Figure 4.3 Example of the result graph

By displaying the relationships of personal attributes in hierarchical, we found that this method allowed us to extract a significant meaning of relationships when consumers disclose their personal attributes. Interpreting the relationships among personal attributes could bring us benefits for many purposes such as:

1. *Reduce overall value of the requested information*

By selecting the requested personal attributes according to the hierarchical level, service providers could possibly eliminate personal attributes that consumers did not feel comfortable with disclosing in the hierarchy. Therefore, service providers could select only the personal attributes that related to their work. When the overall value of requested information decreased, service providers could decrease monetary incentive to consumers.

2. *Increase the number of collected personal attributes*

When service providers had already requested a personal attribute in a higher level of the node, service providers increased the possibility of consumers disclosing their personal attributes from its' child node.

Moreover, this study found that the relationships of in-degree and out-degree among personal attributes could shift depending on the selected participant groups. The graph results from this method allowed change according to any specified group of the consumers, which makes it possible to support personalization. It provides benefits for service providers to analyze their target market. For example, it can be assumed that a service provider wants to offer a monetary incentive to two groups of their target market. In this case, this study compared the relationships among personal attributes between males and females. The first offer was for male consumers, with the second offer being for female consumers. The results for male and female consumers were calculated and shown in Figures 4.4 and 4.5.

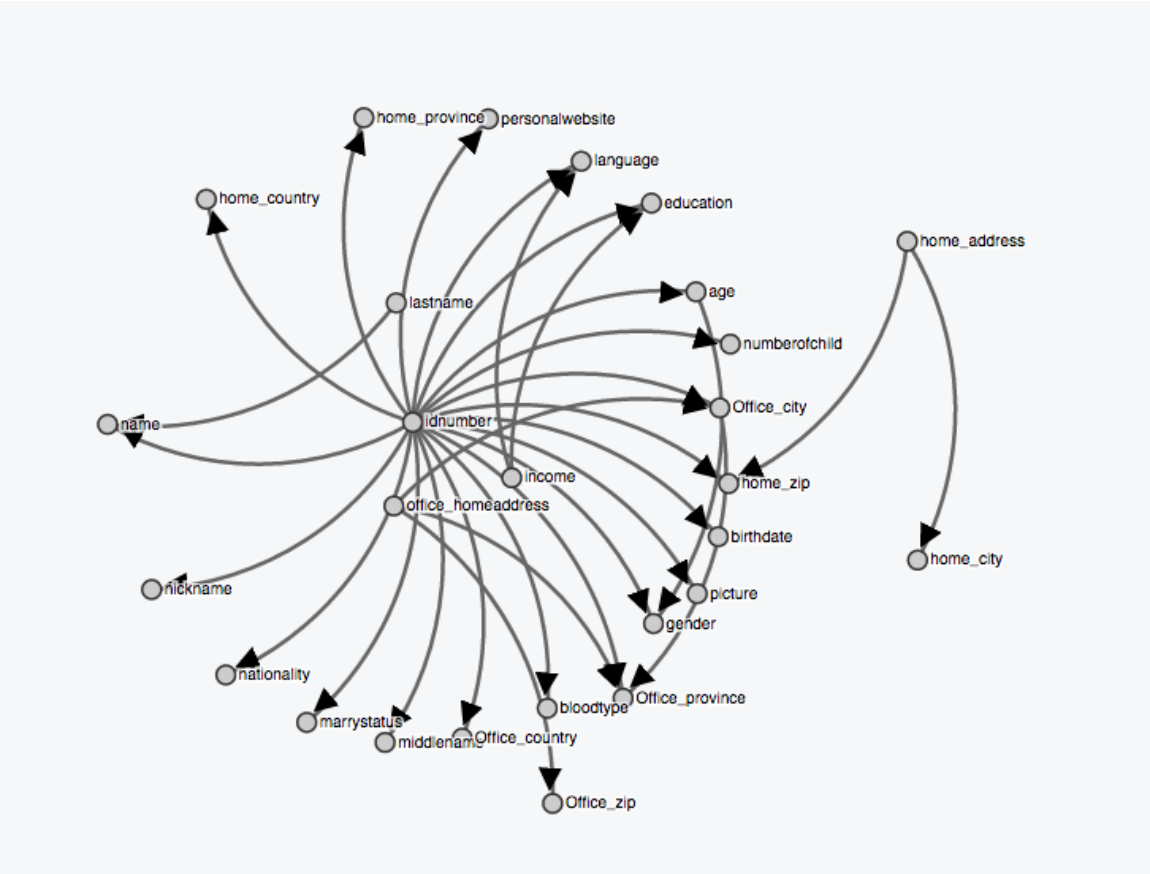


Figure 4.4 Results graph when focused on male consumers

It is clear that the results graphs of Figures 4.4 and 4.5 are different, which show that the estimated value of personal attributes between male and female are different. The relationships value of personal attributes for females was much more complicated than the results graph from males.

For example, *id number* got the highest value for both groups. However, female consumers had a significant concern about their contact information such as *home phone*, *mobile phone,* and *office phone*. This result could be adopted when a service provider requested personal attributes from both groups of consumers. Service providers had to decide carefully when they requested high value personal attributes such as *home phone*, *mobile phone,* and *office phone*. By understanding how each group of consumers viewed their personal information, it is believed that service providers could improve the possibility of collecting personal information from consumers.

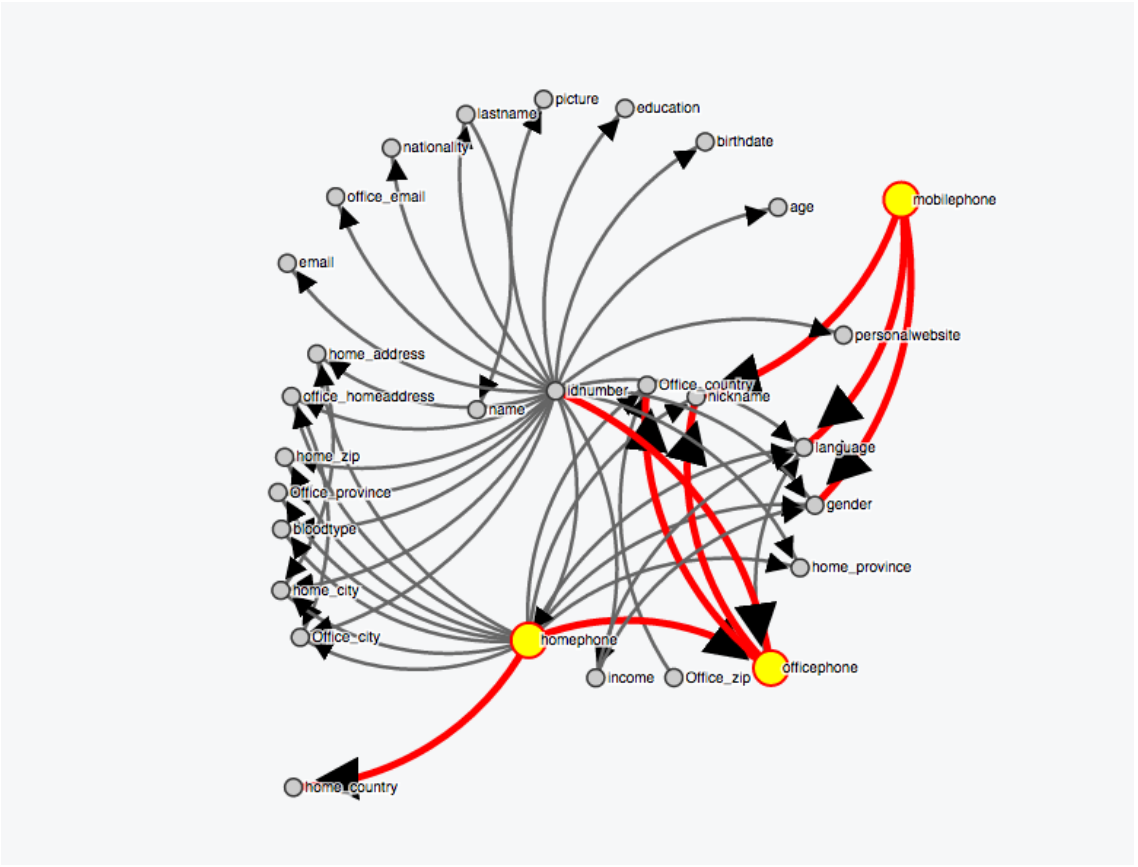


Figure 4.5 Results graph when focused on female consumers

In Figure 4.5, the graph results visually show interesting relationships between personal attributes. The graph shows that the personal attributes that are in the same group are usually linked together. For example, *first name*, *middle name* and *last name* are linked together. Thus, the results from the observation assume that people will protect personal attributes that have semantically similar meanings.

4.2 Method for Personal Information Clustering

At the end of the previous section, the authors offered the assumption that consumers will protect the personal attributes that have semantically similar meaning. This study used a personal attribute clustering method used in a community-detection technique to extract the community structure of disclosure for personal information using the modularity technique.

The technique in [93] aims to maximize the modularity of the clustering result Q, which is defined as follows:

$$Q = \frac{1}{m} \sum_{i,j \in c} \left[weight_{i,j} - \frac{d_i d_j}{m} \right] \quad (2)$$

where m is the number of edges in the network. and c is the set of nodes. $Weight_{i,j}$ is the calculated probability of the edge (i, j) ,and $d_i d_j$ is a multiplication of the degrees of node i and j.

In this study, the results graph was clustered using the above method. It was displayed using different colors in the graph of Figure 4.6. The weight for each edge was the calculation results from the previous section.

The modularity was obtained using the method from the graph at 0.26, which was significantly high considering the fact that the graph was very sparse. This result indicated that the graph had community structure.

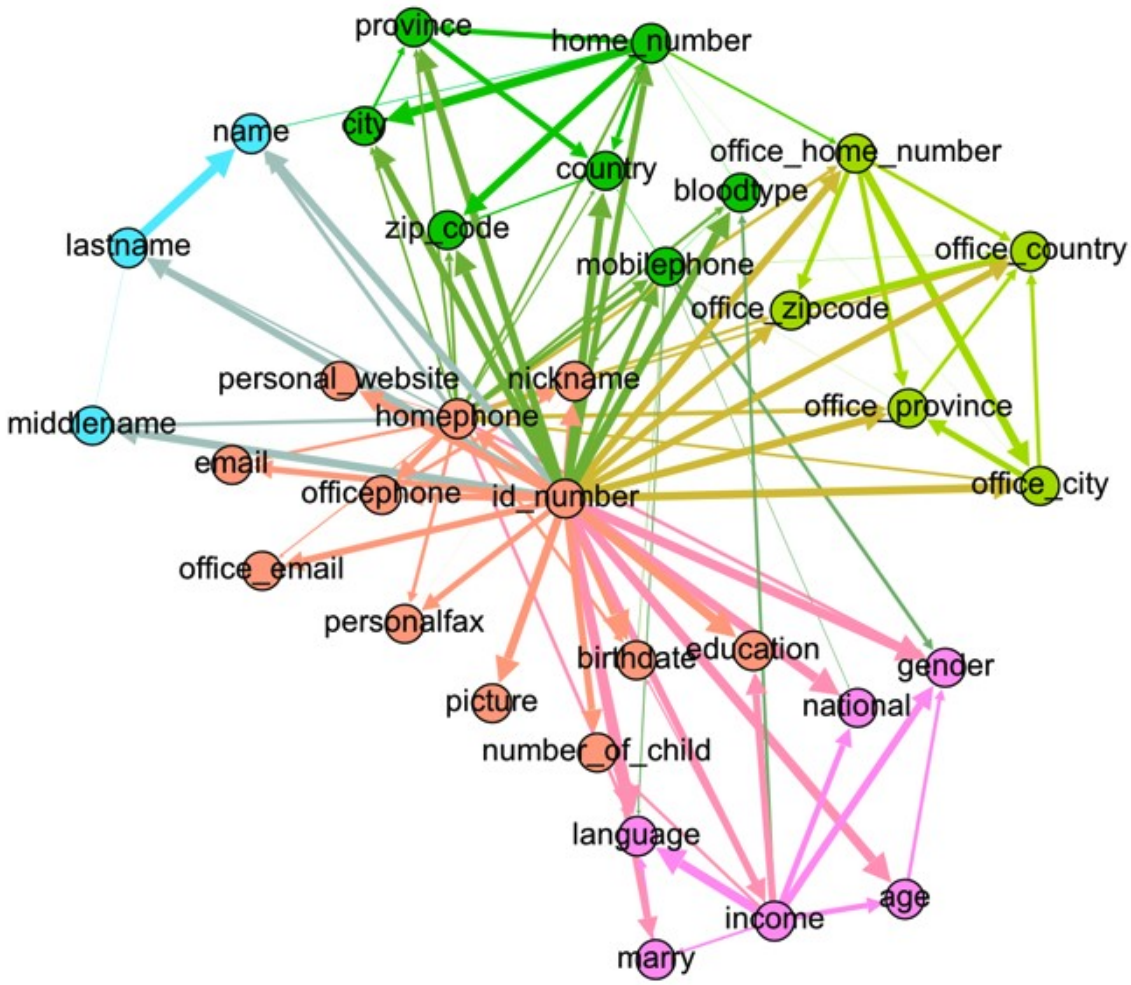


Figure 4.6 Clustering results

To evaluate the assumption that each cluster contained personal attributes that were semantically similar, the authors categorized personal attributes in the graph into five categories, manually by context. Each category contained personal attributes that had similar meaning.

The lists for these categories are as follows:

- Name: Personal attribute related to personal identification by name
- Home address: Personal attribute related to home address
- Office address: Personal attribute related to office address
- Contact information and Personal identifiable information: Personal attribute that is able to identify a particular individual

- Other: other personal attributes not in other categories, such as demographical information

The authors compared the manual classification results with the clustering results in Table 4.1. The results were very similar, with. 84.84% of the personal attributes correctly classified. Precision and recall were calculated for each category of personal information and the results shown. The results showed restricted relationships among the personal attributes in each category.

Table 4.1 Precision and recall of the calculated result

Categories	Manually Classified Result	Calculated Result	Precision	Recall
Name	First name, Last name, Middle name, Nick name	First name, Last name, Middle name	1	0.75
Home address	Home number, City, Province, Country, Zip code	Home number, City, Province, Country, Zip code, Blood Type, Mobile Phone	0.71	1
Office address	Office's home number, City, Province, Country, Zip code	Office's home number, City, Province, Country, Zip code	1	1
Contact information and personal identifiable information	Home phone, Mobile phone, Office phone, Email, Office Email, Personal website, Fax number, Id number, Birth date, Picture	Home phone, Office phone, Email, Office Email, Personal website, Fax number, Id number, Birth date, Picture, Nick name, Education, Number of children	0.75	0.9
Other personal information	Gender, National, Age, Income, Language, Marriage status, Education, Number of children, Blood type	Gender, National, Age, Income, Language, Marriage status	1	0.67

Using the clustering method, the classification was accurate. Moreover, it also showed us some interesting results. We can translate some hidden relationships among personal attributes from the graph. For example, we manually categorized *nickname* under 'name' category, however, it was clustered under 'contact information and personal identifiable information type'. From the result graph, we found that *nickname* had edges that connected to *home phone* and *mobile phone* as demonstrated in Figure 4.2. It was quite interesting that the disclosure of *nickname* attributes did not show the relation to the disclosure of *first name* and *last name,* as we manually categorized it under name category. On the other hand, the clustering result showed us that the disclose of *nickname* was related to *home phone* or *mobile phone* instead. For human, this result of the cluster was quite understandable, since people would told their nickname only to people who were close to them which were people who know their home phone and mobile phone. Therefore, clustering method could help us many hidden relationships among personal attributes. The results show that the classifications decided by the humans were sometimes different from the real action.

4.3 Prototype of Decision Support System for Privacy-Service Trading

This section aimed to verify the assumption that the understanding of value for personal information can improve personal trading activity between service providers and consumers. This case study focused on the service providers' side when offering the value of each personal attribute. The authors implemented a prototype for a decision support system, which aimed to assist service providers when requesting personal information from consumers. The prototype application was called the prototype decision support system for privacy-service trading (DSSPST).

The prototype system was integrated with the proposed method for personal information valuation, as previously described in this chapter. The prototype system was designed to assist service providers in the following aspects:

- *Decrease trading cost*:

When service providers requested personal attributes, regarded by consumers as important information, they must pay more incentives to attract more consumers. Therefore, service providers collected personal attributes, regarded as low importance to consumers, then possibly decreased the trading cost.

- *Understanding the value of personal attributes*:

Service providers usually request personal attributes as they see fit. Consumers may reject this request when they feel uncomfortable about providing personal attributes regarded as important. When a service provider creates a trade request condition offering incentives in exchange for personal attributes, then the value of each personal attribute must be shown.

- *Change the target market*:

Each group of consumers had different ideas concerning the value of their personal attributes. For example, males and females had contrasting views on the importance of their personal information. Service providers need to understand this to create effective trading.

A. Design of the DSSPST

The decision support system for privacy-service trading was designed to visualize the estimated value of each personal attribute. Since this study believes that knowing the value of each personal attribute could help service providers adapt their personal information trading activities, the targeted users for this decision support system were service providers. The user interface design is shown in Figure 4.7. It shows an example using a graph constructed from a specified group of participants. The figure shows nodes for each personal attribute based on the preferences of service providers.

To select the target groups of consumers in the system, each service provider chooses the set of consumers and the set of personal attributes that they are interested in. Personal attribute options in the system include consumer gender, consumer age range, consumer

education level, and consumer occupation. Service providers could select a group of nodes representing user's professional attributes, such as *office address* or *office phone number*.

B. Architecture of the DSSPST

The new system proposed a form of multi-tier architecture. There were three tiers in the system, which consisted of the following:

1) Client Tier: The client tier interacted with service providers via a web browser. Each service provider selected a target market from the provided criteria, including gender, education, and age. They then submitted their criteria to the DSSPST system, which allowed users to view the results on the web browser.
2) Application Tier: The application tier was on the server-side. This contained many modules comprised of identity management, business logic calculation and data collection modules.
3) Data Tier: The data tier included the database server and data management applications. For this study, the database contained consumer data.

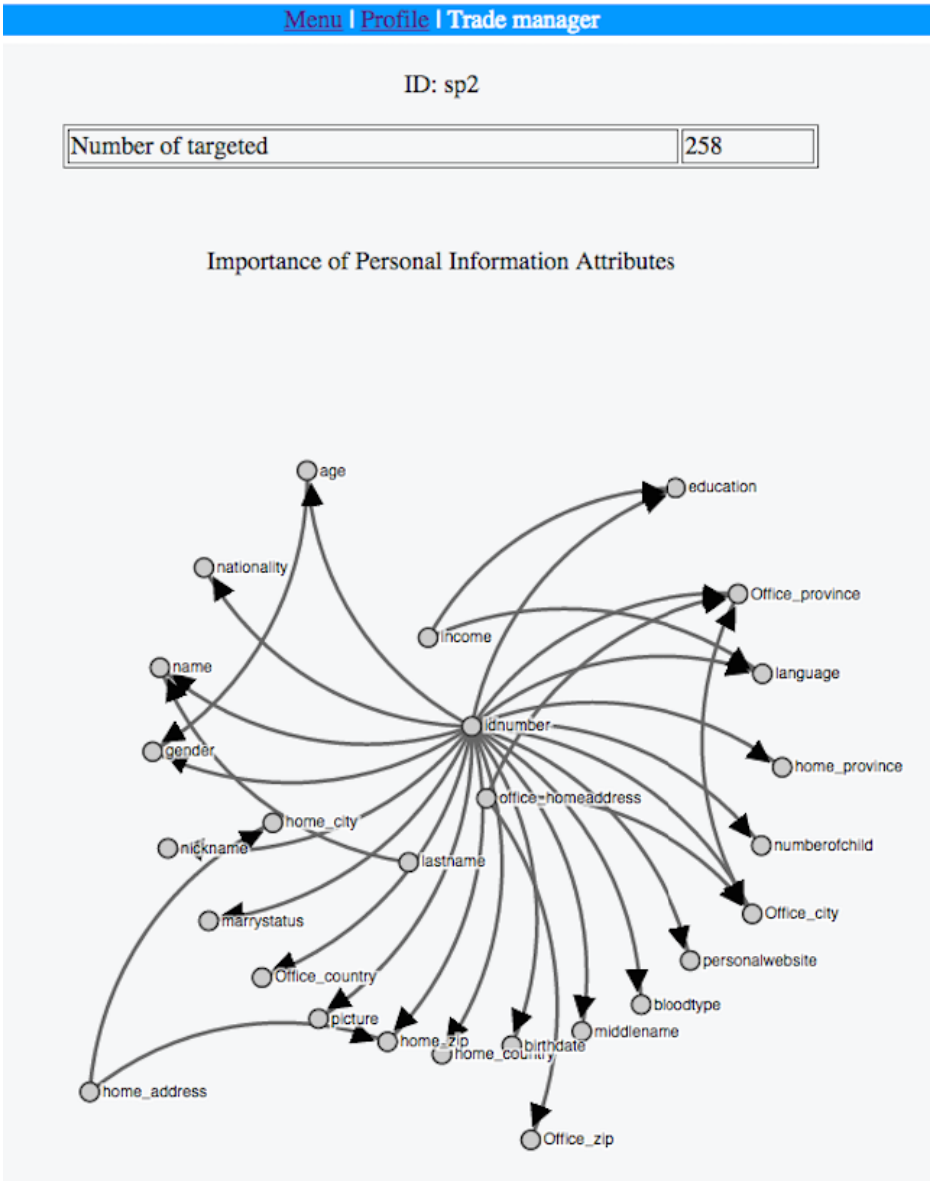


Figure 4.7 Prototype DSSPST

C. Evaluation and Results

Once the prototype DSSPST was developed, the authors invited 14 e-commerce website owners as participants to evaluate the system since the target users of the system were service providers. Firstly, they were asked to create five campaigns to trade between personal information and incentives using their traditional method and without using the new system. Each of them selected five personal attributes for the campaign with different groups of target markets. Secondly, the authors asked them to perform the same task with

support from the prototype DSSPST system. Once both tasks were completed, the authors asked them to complete a questionnaire survey showing satisfaction about the prototype of DSSPST based on a scale from 1 to 5.

The authors found that using the prototype system had an effect to service providers' decisions to adjust their trading design for requesting personal information. It confirmed the study's assumption that understanding the value of each personal attribute could help service providers make better decisions. Service providers changed their decisions when they understood the information of value for each personal attribute from consumers after using the DSSPST. Service providers avoided requesting personal attributes that had high importance in consumers' views. Statistically, the participants were satisfied with the DSSPST result which we had average satisfaction was 4.0 and standard deviation was 0.70.

4.4 Summary

From previous studies, the results from many personal information valuation methods showed that the value of personal information was sensitive to context. It is difficult to estimate the value of personal information in exact numbers. In this chapter, the authors proposed an estimation method for personal information. It used data from a previous chapter and constructed a graph to display the value of each personal attribute. The constructed graph also displayed the relation among personal attributes. It helps for comparison of the importance of personal attributes in consumer's point of view. The results graph can give an advantage to other applications, such as personal information trading system and anonymity improvement.

From the assumption of this study, service providers who understood the value of personal information in consumers' point of view may affect the decisions of service providers when they requested personal information from consumers. This study created a decision support application to verify the above assumption. A prototype decision support system, called DSSPST, was developed for improvement of privacy-service trading. The new decision support system was created by recommending the value of each personal

attribute to service providers. The authors created a prototype of a decision support system, which assists service providers to create an offer for trade between monetary incentive and personal information. The authors selected a group of participants including website owners to use the prototype system. The results showed that service providers changed their decisions when made aware of the value of personal information. They avoided requesting personal attributes which had high value. The evaluation results showed that it was possible to reduce unnecessary requests for personal information. When service providers created a trading condition between personal information and incentives, they usually requested unnecessary personal information. Service providers could avoid requesting important personal attributes if using the DSSPST. The results of the case study confirmed the assumption of this study and explained the initial evolution development of a trading platform between personal information and monetary incentive development.

Lastly, the constructed graph provided an interesting result when clustering personal attributes from the graph. Using the clustering method, the classification was likely accurate, which could imply that people tend to disclose personal attributes that are semantically similar. Moreover, the graph also showed hidden relationships among personal attributes from the graph. Presently, service providers have more chance to collect personal information, although the number of personal attributes in some applications, such as SNS, can be very large for clustering personal information manually. The authors cannot evaluate the importance of those thousand personal attributes manually. The proposed method potentially helps service providers to classify personal attributes collected from their consumers.

CHAPTER 5

PRIVACY DISCLOSE ADAPTION FOR TRADING PLATFORM

The results in the previous chapter showed that understanding the value of personal information can provide advantages for both service providers and consumers. Although the graph obtained in Chapter 4 was robust against the evaluation context, it was relatively hard to use. The previous graph in Chapter 4 can be an effective visualization tool for showing the comparison of value between personal attributes, but it is still difficult to understand the meaning, especially when there are many nodes to judge and needed for ranking. This study improves previous work regarding the valuation of personal information disclosure for use in the trading platform. It proposes a new requesting personal attributes approach, which possibly adapts consumers' personal information disclosure behavior and aims to increase the disclosure of personal information without increasing monetary incentive. The proposed method is used in the evaluations, which compares the disclosure of personal information results from consumers. After the

evaluation is completed, the results show that the new approach could increase the disclosure of consumers' personal information.

5.1 Overview

The trading mechanism for trading platform was discussed in Chapter 2. The trading platform was designed as a medium between personal information and monetary incentive. The common steps are as follows: firstly, the service provider prepares an incentive, requesting the consumer to disclose a particular personal attribute. Secondly, the consumer receives the offer and then makes a decision. Lastly, he/she will receive the incentive on condition of agreement to disclose the personal attribute.

The authors proposed a method that used the graph to show the relation of disclosure between each personal attribute for a specified group of consumers in Chapter 4. The graph provided useful information and was able to compare personal attributes visually. Even though the graph can be an effective visualization tool for showing the comparison of value for each personal attribute visually, it is still difficult to use in an application such as a trading platform. With many personal attributes, the constructed graph was very large and proved complicated to translate meaning. To measure the value of one personal attribute and compare it with another, the trading platform needs a better method. This chapter proposes a new numerical value called the *Value of Unwillingness to Disclose (VD)*.

5.2 Development of Valuation of Unwillingness to Disclose

This chapter aims to improve previous work regarding the valuation of personal information disclosure for use in a trading platform with the above respects. This study converts the graph results in a previous chapter into a new tree. Then, the tree is used to calculate new numerical values called the *Value of Unwillingness to Disclose (VD)*. *VD* is calculated from the probability that participants protected personal attributes, personal attribute that contains high *VD* means consumers want to protect this personal attribute more than other personal attributes that contain lower *VD*.

The authors calculated the probability that participants protected their personal attributes when they had already protected another attribute. Then, the authors constructed a direct graph in Figure 4.1. Next, the edges of the graphs were eliminated that contained low probability and focused on the pairs of personal attributes with strong relationships. The results graph in Figure 4.2 showed that consumers gave value to each personal attribute differently. It also showed that consumers disclose their personal attributes in a hierarchy. For example, they will disclose their first name before their last name. A tree structure was adopted to show this obviously.

The graph obtained in the previous chapter is converted to trees using topological sorting [94]. This obtained a tree with the root nodes as the most important personal attributes and leaves are the less important personal attributes. A node with a large outdegree in the graph is an important node, so the topological sort algorithm starts at the edge with the largest outdegree. Nodes with large indegrees in the graph are usually not important. Since they are visited after all nodes that have edges to them, they tend to be leaves that are far from the root node. A topological ordering technique is employed to convert the graph into a new tree for better comprehension.

The transformation steps can be described as follows:

a) Find a new root node. The root node is a node that has no indegree, but the highest outer degree. The root node then becomes a parent node.

b) Select the child nodes that do not have any outer degree. Then, connect them as a child from the parent node.

c) Select a child node which has only one indegree, which is the parent node. Then, connect the selected child node to its parent. The new child node becomes a new parent node.

d) Repeat Steps b to d until completing a new arrangement for all nodes.

The graph in Figure 4.2 is transformed into the tree shown in Figure 5.1. This resulting tree is easier to understand than the previous graph. Personal information which is most important to customers and the hierarchy of personal information disclosure can now be seen more clearly. However, this can still become complicated with large datasets and high numbers of nodes.

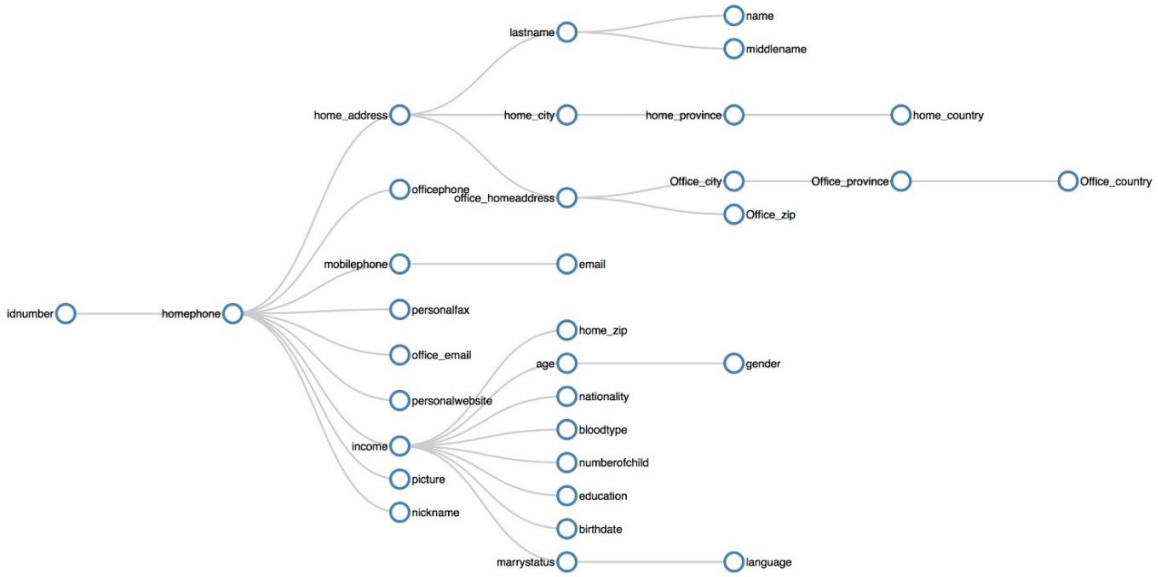


Figure 5.1 The results tree graph

Let A be a node in the graph and a personal attribute. Let R be a root of the result tree. If A=R, then

$$VD_A := 1. \tag{3}$$

If not, assume that $A_n := A$ and the path from R to A in the tree is $[(A_0 := R, A_1), (A_1, A_2),..,(A_{n-1}, A_n)]$ then

$$VD_A := \prod_{i=1}^{n} W_i , \tag{4}$$

where

$$W_i = P(A_i \mid A_{i-1}) = \frac{P(A_i \bigcap}{P(A_{i-1})} \quad (5)$$

By the argument in Chapter 4, we know that W_i represents the relative importance of A_i comparted to A_{i-1}. The value of VD_A is then a value representing the relative importance of personal attribute A compared to the most important personal attribute R. The authors do not directly assign $P(A \bigcap$?) to VD_A, because A and R do not have a strong relationship when there is no edge between A and R and the authors strongly believe that relative importance W_i should be calculated only from two personal attributes with a strong relationship.

Since *VD* was calculated from the probability that participants protected personal attributes, personal attribute that contains high *VD* means consumers want to protect this personal attribute more than other personal attributes that contain lower *VD*. The results obtain from this calculation are shown in Table 5.1.

Table 5.1 Value of Unwillingness to Disclose

Number	Attribute	VD
1	National ID Number	1
2	Home phone	0.950
3	House No.	0.865
4	Office phone	0.858
5	Mobile phone	0.892
6	Fax (Personal)	0.876
7	Office email	0.744
8	Personal website	0.697
9	Income	0.880
10	Picture	0.805
11	Nickname	0.221

12	Last name	0.737
13	Home city	0.625
14	Office No.	0.759
15	Email	0.616
16	Home zip	0.524
17	Age	0.384
18	Nationality	0.219
19	Blood type	0.410
20	Number of children	0.477
21	Education level	0.341
22	Birthdate	0.678
23	Marital status	0.437
24	First name	0.587
25	Middle name	0.509
26	Home province	0.448
27	Office city	0.464
28	Office zip	0.441
29	Gender	0.190
30	Language	0.183
31	Home country	0.296
32	Office province	0.361
33	Office country	0.261

5.3 Experiment

A web application was created to simulate trading situations using the trading platform. The web application was developed using PHP language and hosted on a private server. In this study, the participants were Thai Internet users. The participants were invited to register to use the web application. Then, the system displayed monetary incentives, which were fixed as gift vouchers worth 100 baht. The system showed a condition to participants for each person to receive the maximum value of the incentive provided when they

disclosed all personal attributes. The value of the incentive was reduced incrementally depending on which personal attributes they declined to disclose.

Next, the participants were asked whether they would provide the displayed personal attributes. There were two options for them, either disclose or reject. Regardless of the selected choice, the application asked the same question for the next personal attribute. Participants answered the questions until they arrived at a finish page. There were 33 personal attributes questioned in this application, which was the same number of personal attributes set for calculated *VD*. Figure 5.2 shows a screen shot of the web application when asking the questions to participants.

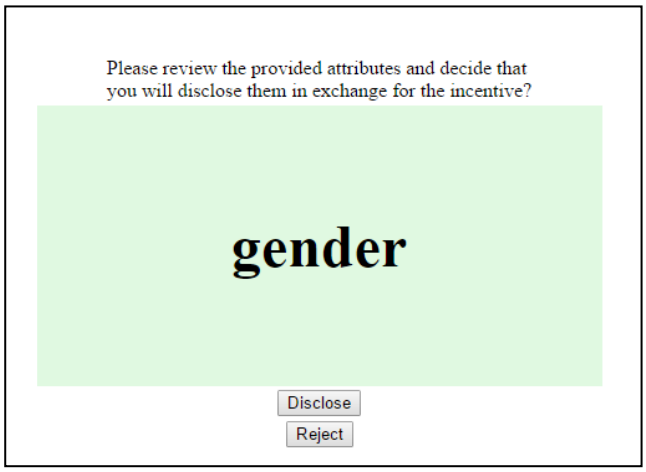


Figure 5.2 Screenshot of the web application asking a disclosure question

The order of personal attributes questioned in this study aims to improve trading between personal information and monetary incentives. This study separated the participants into two groups. Each group used the web application that questioned a different set of personal attributes.

The two sets of questions were called "*top-down approach*" and "*bottom up approach*". The top-down approach used the *pre-order traversal* [94] to order the personal

attribute questions from the top of the tree downward into all child nodes. In other words, the ordering travelled from the highest *VD* node at the top to the lowest *VD* node at the lowest level of the tree.

The bottom-up approach used the *post-order traversals* [94] to order personal attribute questions from the leaves of the tree at the lowest level of the tree, with ordering travelling up into the root nodes. The ordering travelled from the lowest *VD* node to the highest *VD* node.

When there was more than one node in a level, the ordering approach selected the node that had the highest *VD* in the top-down approach or selected the node that had the lowest *VD* in the bottom-up approach.

For instance, Figure 5.3 is a tree containing 7 personal attributes, represented by nodes 1 to 7. The tree has 0 to 3 levels. The root node is node 1 on level 0. Node 2 is on level 1. Nodes 3 and 4 are on level 2. Nodes 5, 6 and 7 are on level 3. The top-down approach selects a set of personal attributes as 1, 2, 3, 4, 5, 6 and 7. On the other hand, the bottom-up approach start the order from the lowest *VD* on the lowest level. The bottom-up approach selects a set of personal attribute as 7, 6, 5, 4, 3, 2 and 1.

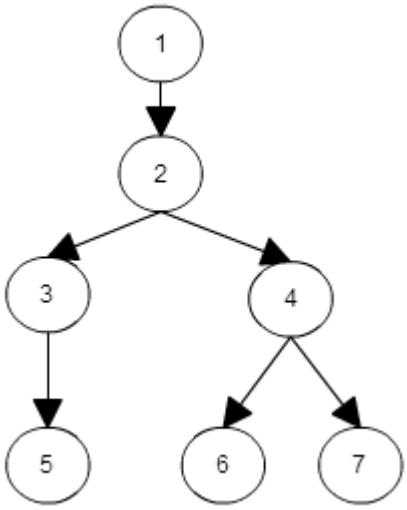


Figure 5.3 Example of a tree for the ordering approach

5.4 Experiment Results

From previous studies, the authors believe that service providers can gain more benefits when they understand consumer value for disclosure of personal information. In this study, two consecutive experiments were conducted.

In this experiment, the authors invited 100 participants to use the web application. The invited group included internet users in Thailand. Participants from Thailand were selected because the authors had collected, used, constructed, and calculated the tree and *VD* from this group of users in Thailand. A different group of users may affect judgment in disclosing personal attributes. For example, National ID number has a high *VD* in Thailand, but it may not affect consumers in other countries the same way. The participants were separated into two groups for the first experiment. Each group completed a different approach, either from top-down or bottom-up approaches.

The experiments were completed by participants and their answers collected. The authors compared the results with the total *VD* when all personal attributes were disclosed of 18.735. The average of *VD* from each group of participants is shown in Table 5.2. When the top-down approach was used, the average of total *VD* was 11.263, which was 60.12%. When the bottom-up approach was used, the average of total *VD* was 9.5254, which was 50.84%. To test our hypothesis in this study, the authors used p–value at the conventional criteria of 0.05 as a threshold.

To test Hypothesis H1: *The top-down approach is better than the bottom-up approach*, the *p*-value was calculated using Welch's *t*-test [95]. The *p*-value obtained from the calculation was 0.0591. Although the value was still higher than the conventional criteria of 0.05, we believe that the value was small enough to conclude that the top-down approach was better than the bottom-up approach.

The authors subsequently conducted the second experiment. The results from the top-down approach were used as a baseline in the second experiment, which aimed to

improve consumers' disclosure of personal information. The authors invited more 60 participants for this experiment and adapted the order of personal attributes in the web application according to their profiles. In this study, the authors selected gender as a criterion because it was found that there were differences in their disclosure. New trees were constructed for males and females, with the ordering of personal attributes rearranged in the web application.

For the second experiment, ordering was enhanced from the last top-down approach using the demographic data of the consumers. For example, an order was constructed from the 258 female participants in subsection 2.1 if the consumer participating in the survey was female. The results obtained following the improvement are shown in Table 5.2. This technique is termed *the adaptive approach*. When the adaptive approach was used, the average of total *VD* was 12.393, which was 66.15%.

Table 5.2 Results of Top-Down, Bottom-Up and Adaptive Approaches

Approach	*Average*	*SD*	*#Participants*
Top-down	11.263	4.533	49
Bottom-up	9.525	4.563	51
Top-down with the adaptive approach	12.393	4.192	60

To test Hypothesis H2: *the top-down adaptive approach is better than the top-down approach*, the *p*-value was calculated using Welch's *t*-test [95]. Unfortunately, the *p*-value obtained from the calculation was 0.180. Although the top-down adaptive approach had a significantly higher average disclosure, the authors could not conclude that the statistic was significant.

On the other hand, the *p*-value for Hypothesis H3: *the top-down adaptive approach is better than the bottom-up approach* was 0.0008. While it was not very clear that the top-down approach improved the bottom-up approach by the *p*-value of H1, the improved version of the top-down approach clearly improved the bottom-up technique. The result from adaptive approach is 6.03% better than Top-down approach.

The percentages of personal attribute disclosure for participants were calculated and are shown in Figure 5.4. The graph displayed the difference in results of consumer disclosure for each personal attribute between the top-down approach, bottom-up approach, and enhanced approach. The results of the top-down approach show that participants disclosed their personal attributes easily when the web application started questioning from high *VD* attributes to low *VD* attributes. Participants may disclose personal attributes with low *VD* easily in both the top-down approach and bottom-up approach, but the percentage to disclose personal attributes with a high *VD* value significantly decreased in the bottom-up approach.

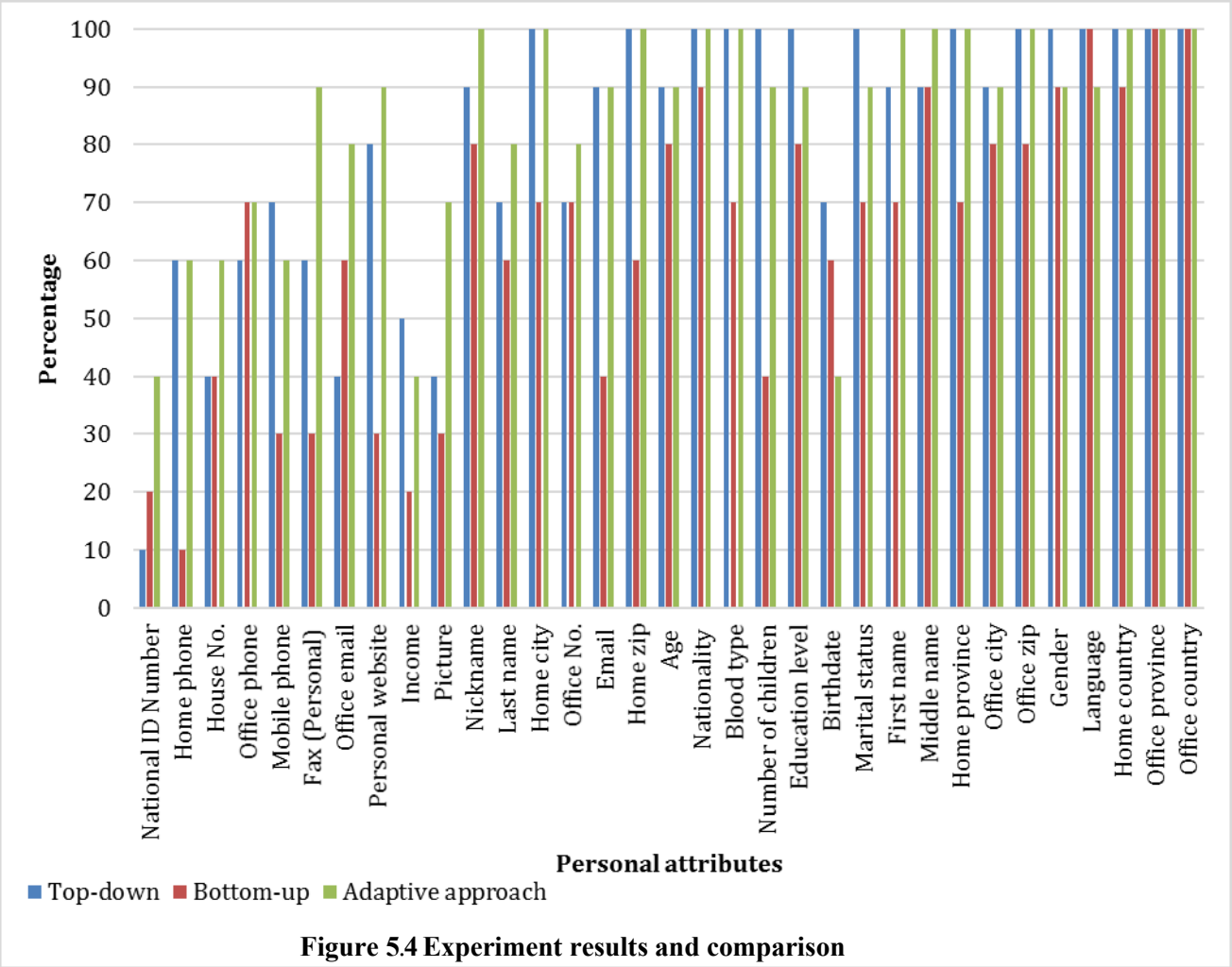


Figure 5.4 Experiment results and comparison

The adaptive approach of this study has been used to improve the top-down approach. The results of the adaptive approach showed that the disclosure of personal attributes for participants can be increased. From the graph 5.4, the disclosure of personal attributes for participants increased steadily for low *VD* attributes because the top-down approach result is already effective for the disclosure approach. In addition, the new approach results in a significant increase in high *VD* personal attributes. The results of the adaptive approach support this study's assumptions that consumers disclose their personal attributes in a hierarchical form and that the personal attributes have semantic similarity between them.

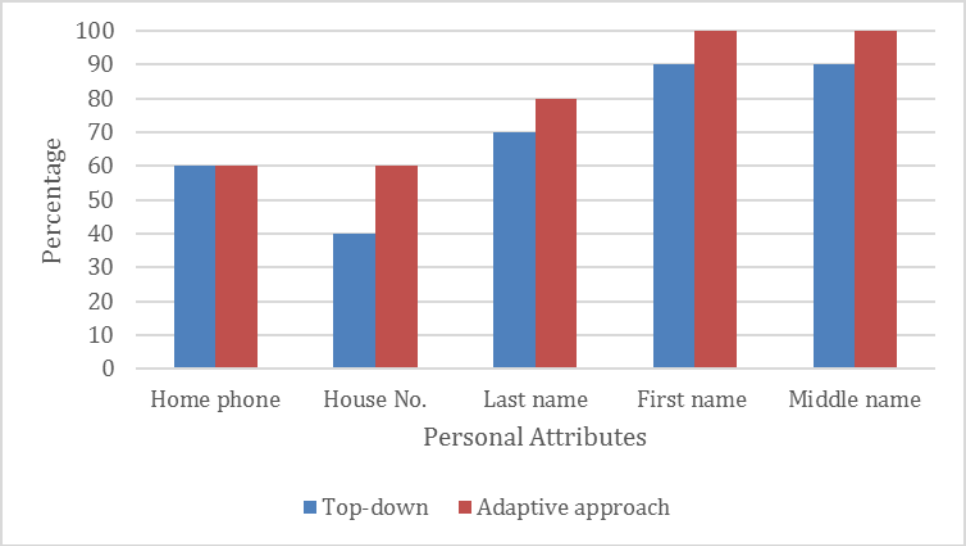


Figure 5.5 Example of the comparison result

Figure 5.5 is an example of the disclosure results in the percentage form from a set of personal attributes. These personal attributes have semantic similarity and hierarchy in the graph shown in Figure 5.1. From calculations in this study, home phone has the highest position in the tree and participants disclosed personal attributes in the form of a hierarchy under the home phone node until reaching the leaf node, comprised of first name and middle name. The results in Figure 5.5 show that the disclosure ratio increased when using the adaptive approach.

5.5 Summary

From previous studies, it was found that service providers can gain more benefits when they understand consumers' disclosure of personal information. This study used knowledge from previous chapters to improve the trading activities between service providers' monetary incentives and consumers' decisions to disclose their personal information when receiving service providers' offers and incentives.

From the previous chapter, a valuation method was proposed to compare personal attributes using a graph. In this chapter, the authors enhanced the method by using a technique to transform the graph into a new tree with numerical values called the *Value of Unwillingness to Disclose* (*VD*). Then, the authors calculated and assigned *VD* value to all personal attributes in the tree. The tree was analyzed and sets of personal attributes were

created with each set of personal attributes ordered differently according to their *VD*. Next, an application was created to simulate trading situations using the trading platform.

The results indicate that consumers tended to provide more personal information when the questions are ordered from the most important personal attribute to the least important. This improvement is more significant when the order was obtained from survey data on participants with the same demographic grouping.

CHAPTER 6

CONCLUSION AND DISCUSSION

At the beginning of this thesis, we introduced the research problems of personal information collection, the thesis objective, and devised alternative methods to improve the collection of personal information for trading activities between personal information of consumers and monetary incentives of service providers. In the final chapter, we conclude overall discovery of this thesis and make recommendations for the future works.

6.1 Conclusion and Discussion

This thesis aimed to investigate the exchange mechanism between monetary incentives from service providers and personal information from consumers. From study of earlier works, we understood that personal information is an important asset for consumers as it can be bought, sold, and exchanged. Researchers and businesses adopted the ideas of buying and selling personal information. The trading of personal information is still

considered a complicated activity. The problem of the exchange mechanism is not related to limitations of technology, since current technology can easily support these trading activities. Numerous tools and technology exist to provide channels for communications between consumers and service providers.

Devices connected to the internet owned by consumers create and send large amounts of personal information over the Internet every day. This situation allows service providers to easily collect, store, and use consumers' personal information. Consumer currently have more concern and awareness of their privacy [96]. However, the major problem is how service providers request this personal information, because there is a difference in point of view between service providers and consumers about the value of this information. Service providers require large amounts of personal information while limiting monetary incentives offered to consumers. On the other hand, consumers are reluctant to provide their personal attributes, and demand more monetary incentive from service providers for doing so. There is no solution for this situation.

This thesis focused on service providers who ask for personal information from consumers. We aimed to increase the quantity of personal attributes which service providers can collect from their consumers. To improve the current trading situation, we looked at the problem of valuing personal information. Many studies of this used different methods for calculating the value of personal attribute, however, the results are diverse and unreliable. This realisation led us to find a new method of estimating the value of personal information to support the exchange mechanism between monetary incentive offered by service providers and the valuable personal information of consumers.

The starting point for this thesis was an investigation of the demand of personal information by service providers and disclosure level of personal information by consumers. Results of this comparison show differing points of view about the value of personal information between service providers and consumers. Valuation from service providers have significantly larger variation than those from consumers. Service providers have personal attributes with significantly larger valuation than the others. Each type of service provider has different demand for personal attributes. By contrast, consumers give more similar valuation for most of personal attributes.

This thesis proposes a new method for personal information valuation for use in trading activities. We studied methods proposed in other studies, but found their methods difficult to apply practically. Information from these works showed that people judge the value of their personal information differently. Most works based the value of personal

information only on tangible benefits such as market price. However, many studies showed different results when they calculated the value of each personal attribute, showing that people are prepared to disclose personal information for both tangible and intangible benefits. From this aspect, we proposed an estimation method for personal information using consumers' attitudes towards the disclosure of each personal attribute. We use probability and graph techniques to estimate the value of personal information. The proposed method gives advantages suitable for trading activities, because it is easy to adapt results for different groups of consumers or participants. It also shows estimated values of personal information in a hierarchy, which lends itself easily to support selection in trading mechanisms.

Our method involves constructing a directed graph that shows the relationships between personal attributes disclosure. This allows us to understand the disclosure mechanism for personal attributes from the point of view of consumers. Moreover, the resulting graph shows that personal attributes cluster rationally.

At this point, we could compare two personal attributes using the graph. However, graphs showing many personal attributes were still too complex for the comparison. So, the graph had been evolved into a new tree which clearly displays the hierarchy of relationships of personal attributes. Then, we proposed a method to calculate the value of unwillingness to disclose each personal attribute, which we called *Value of Unwillingness to Disclose* (*VD*). From the constructed tree, we ranked the value of each personal attribute from consumers' point of view. This also shows the dependency between different personal attributes.

Furthermore, this study uses the proposed valuation method for increasing the performance of trade between personal information and incentives. The trading application was built to find a better method for requesting personal information from consumers. Firstly, we used our tree to create two sets of personal attributes with a fixed monetary incentive. Each set of personal attributes was arranged differently using its *VD*. The first set of personal attributes were organized from the highest *VD* to the lowest *VD,* which we called the *top-down* approach, and the second set of personal attributes were arranged from the lowest *VD* to the highest *VD,* which we called the *bottom-up* approach. Then, we compared the results of participant disclosure.

Results are significantly different between the *top-down* and the *bottom-up* approach. Participants disclose more personal attributes when we arranged personal attributes using the *top-down* approach. Then, we improved the disclosure of personal

information by adapting the constructed tree to suit the participants' profiles. For this thesis, we constructed the tree by participant gender, then created a new set of personal attributes which we called the *adaptive* approach. Results show that the average outcome of the *adaptive* approach is better than that achieved using the *top-down* approach.

This result shows not only that monetary incentives can encourage consumers to disclose personal information, but also that the requesting ordering, as well as the environment of the trading activity affect a consumer's decision to provide personal attributes to service providers.

The result of this study can be compared with another related study of the authors on the ordering of questions, which aimed to improve the motivation of survey participants [97] which organized an experiment providing two sets of questionnaires, each of which asked for personal information, but in differing order. The valuation of each personal attribute was calculated using the proposed technique of the authors. The first set of questionnaires asked about personal information from high value to low value. The second set of questionnaires started from low value of personal information to those with high value. The result shows that participants who received the second set of questionnaire agree to submit a higher percentage of personal information. Results of previous research were different to this thesis. This result showed that the environment in which personal information is requested can affect the quantity of personal information gathered. Using the survey form, participants could review requested personal attributes then select the personal attributes which they were comfortable disclosing.

However, the simulation of the trading platform did not provide any chance for them to review the set of personal attributes requested. When participants received an offer of a monetary incentive, they had to decide without knowing anything about the next requested personal attributes, and they could not change their decision once made. Some participants commented that they wished to change their decision after seeing the next requested personal attribute. This discovery shows the effectiveness of our method, and that it may extend to other areas of related study.

Moreover, the results from the proposed method for estimating the value of each personal attributes clearly show that people treat each personal attribute differently. This idea may be useful in other related situations such as marketing and planning strategies that need to acquire personal information. The results from this study also can provide benefits for service providers who want to exchange incentives with consumer' personal information. Especially service providers who have variety groups of consumers.

Additionally, the usages of the created knowledge possibly extend to other activities such as in Thailand's Personal Data Protection Act. Currently, Thailand does not have any specific statutory law governing personal data protection [98] ; however, the government is in the process of drafting the Personal Data Protection Act [99, 100, 101]. The findings from this study including personal attributes clustering, consumers' attitudes toward personal information, and the ordering of personal information may be useful for organizations, and also relevant to the drafting of this Act in the areas of personal information categorization and personal data inquiry.

6.2 Limitations and Future Directions

The intention of this thesis is to study the trading mechanism which focuses on trading between personal information and monetary incentive. Although privacy trading is a common activity, in which people trade their privacy with everyday life activities, the decision to provide personal information is still complex because everyone evaluates personal information differently.

Even though this study proposed a general model that can be adapted to many situations and many group of consumers, there were limitations related to the participants in the study. The study had a time limit, and focused on a group of Thai Internet consumers as the sample.

Therefore, this study was performed on a limited scale, it could have been improved by using a larger number of people. Different groups of consumers may produce difference results for personal information disclosure. From the fact that people in each country grow up in difference environments, cultures, and local laws, we believe that different groups of international participants would provide interesting results attributable to their different decision-making mechanisms.

APPENDIX A

Questionnaire Form

This appendix displays the questionnaire form which used in Chapter 3. This study used the online form [102] which is a free tool on Google website.

1. Part 1: Demographical Questions.

Edit this forr

แบบสอบถามเพื่อการวิจัย การเปิดเผยข้อมูลส่วนตัวออนไลน์

แบบสอบถามชุดนี้เป็นส่วนหนึ่งของโครงการปริญญาเอก The Graduate University for Advanced Studies［SOKENDAI］จัดทำขึ้นโดยมีวัตถุประสงค์ เพื่อศึกษาความสำคัญของข้อมูลส่วนบุคคลในมุมมองของผู้บริโภค

แบบสอบถามนี้มีทั้งหมด 4 ตอน ซึ่งแต่ละตอนได้ระบุคำแนะนำในการตอบไว้เรียบร้อยแล้ว ขอความกรุณาโปรดตอบคำถามตามความเป็นจริงมากที่สุด ทั้งนี้คำตอบและข้อมูลของท่านจะถูกเก็บรักษา และนำไปใช้ประโยชน์ในเชิงวิชาการเพื่อการศึกษาวิจัยในครั้งนี้เท่านั้น

* Required

Gender (เพศ) *

- ชาย
- หญิง

Age range (ช่วงอายุ) *

- 15-20
- 21-30
- 31-40
- 41-60
- > 60

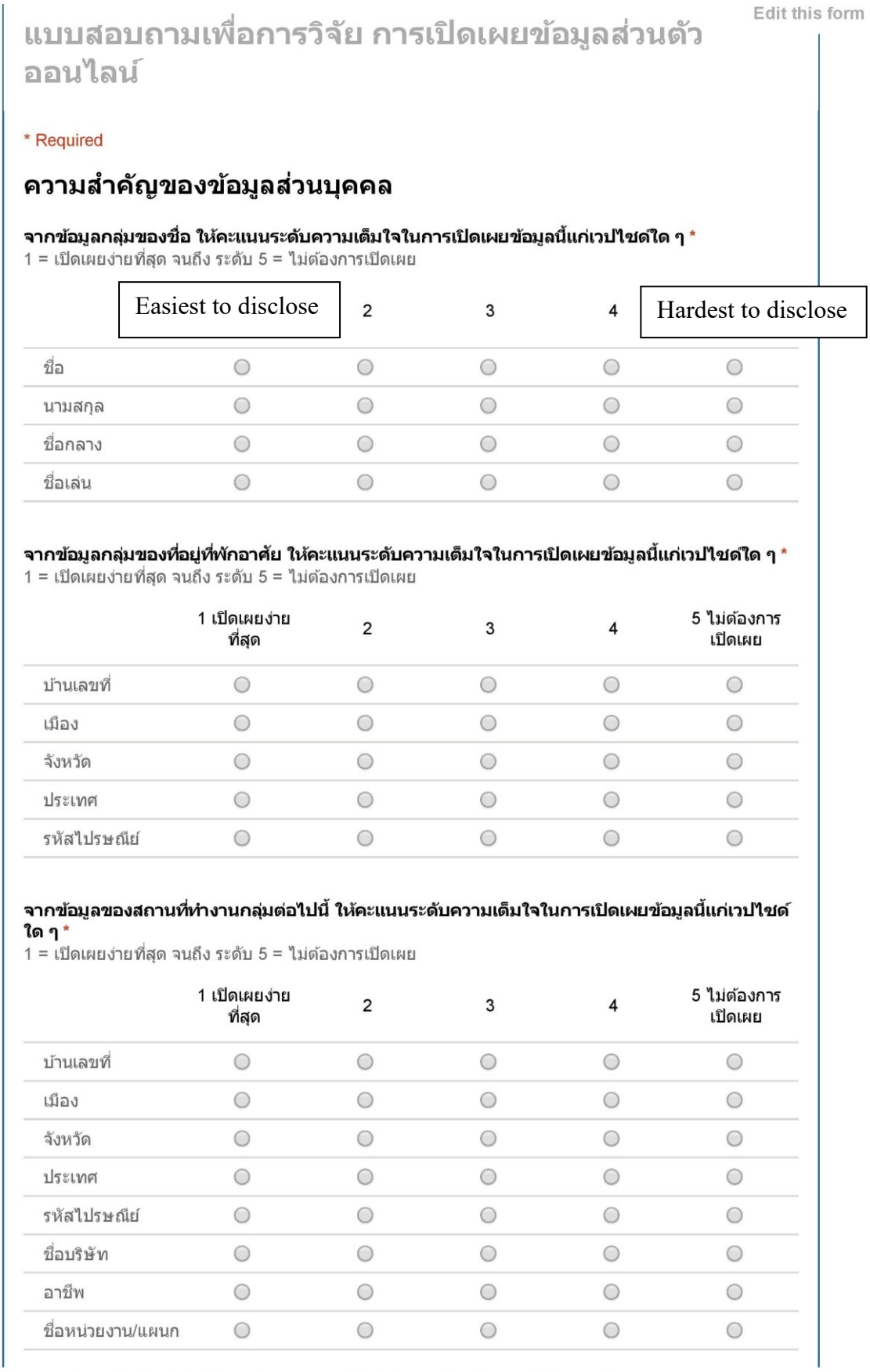
Highest level of education (ระดับการศึกษาสูงสุด) *

- มัธยมศึกษา High school หรือ เทียบเท่า
- วิทยาลัย College หรือ เทียบเท่า
- Bachelor's degree ปริญญาตรี
- Graduate degree สูงกว่าปริญญาตรี

Career (อาชีพ) *

- นักเรียน/นักศึกษา
- ค้าขาย/ธุรกิจส่วนตัว
- พนักงานบริษัท/องค์กรเอกชน
- รับราชการ/เจ้าหน้าที่รัฐ
- รัฐวิสาหกิจ
- ว่างงาน/แม่บ้าน/เกษียรอายุ
- อื่น ๆ

2. Part 2: Comfortable level when participants disclose each personal attribute questions



Edit this form

แบบสอบถามเพื่อการวิจัย การเปิดเผยข้อมูลส่วนตัวออนไลน์

* Required

ความสำคัญของข้อมูลส่วนบุคคล

จากข้อมูลกลุ่มของชื่อ ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *

1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	Easiest to disclose	2	3	4	Hardest to disclose
ชื่อ					
นามสกุล					
ชื่อกลาง					
ชื่อเล่น					

จากข้อมูลกลุ่มของที่อยู่ที่พักอาศัย ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *

1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
บ้านเลขที่					
เมือง					
จังหวัด					
ประเทศ					
รหัสไปรษณีย์					

จากข้อมูลของสถานที่ทำงานกลุ่มต่อไปนี้ ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *

1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
บ้านเลขที่					
เมือง					
จังหวัด					
ประเทศ					
รหัสไปรษณีย์					
ชื่อบริษัท					
อาชีพ					
ชื่อหน่วยงาน/แผนก					

จากข้อมูลกลุ่มของเบอร์โทรศัพท์ ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *
1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
เบอร์โทรศัพท์บ้าน	○	○	○	○	○
เบอร์โทรศัพท์ที่ทำงาน	○	○	○	○	○
เบอร์โทรศัพท์มือถือ	○	○	○	○	○
หมายเลขโทรสารส่วนตัว	○	○	○	○	○

จากข้อมูลของการติดต่อออนไลน์ ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *
1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
อีเมล์ส่วนตัว	○	○	○	○	○
อีเมล์สถานที่ทำงาน	○	○	○	○	○
เวปไซต์ส่วนตัว	○	○	○	○	○

จากข้อมูลของข้อมูลพื้นฐาน ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *
1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
เพศ	○	○	○	○	○
อายุ	○	○	○	○	○
เชื้อชาติ	○	○	○	○	○
ภาษา	○	○	○	○	○
สถานะการแต่งงาน	○	○	○	○	○
กรุ๊ปเลือด	○	○	○	○	○
จำนวนบุตร	○	○	○	○	○
ระดับการศึกษา	○	○	○	○	○
GPS Location สถานที่ปัจจุบัน	○	○	○	○	○

จากข้อมูลของข้อมูลระบุตัวตน ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *
1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
วันเดือนปีเกิด	○	○	○	○	○
หมายเลขบัตรประจำตัวประชาชน	○	○	○	○	○

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
รายได้	○	○	○	○	○
รูปถ่ายส่วนตัว	○	○	○	○	○

จากข้อมูลของข้อมูลความชอบ และ งานอดิเรก ให้คะแนนระดับความเต็มใจในการเปิดเผยข้อมูลนี้แก่เวปไซต์ใด ๆ *

1 = เปิดเผยง่ายที่สุด จนถึง ระดับ 5 = ไม่ต้องการเปิดเผย

	1 เปิดเผยง่ายที่สุด	2	3	4	5 ไม่ต้องการเปิดเผย
งานอดิเรก	○	○	○	○	○
รายการโทรทัศน์ที่ชอบ	○	○	○	○	○
หนังสือที่ชอบ	○	○	○	○	○
ภาพยนต์ที่ชอบ	○	○	○	○	○

« Back Submit

Never submit passwords through Google Forms.

100%: You made it.

APPENDIX B

Related Publications

1. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara, "Privacy Disclosure Adaptation for Trading between Personal attributes and incentives", Journal of Information Processing, Vol.25 No.1 (Jan. 2017), page 1-10.
2. Ake Osothongs and Noboru Sonehara. "A Proposal of Personal Information Trading Platform (PIT): A Fair Trading between Personal Information and Incentives", International Conference on Digital Information and Communication Technology and its Applications (DICTAP 2014), page 269-274, IEEE, 2014.
3. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara. "Evaluating the importance of personal information attributes using graph mining technique", International Conference on Ubiquitous Information Management and Communication (IMCOM 2015), 8 pages, ACM, 2015.
4. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara. "A Prototype Decision Support System for Privacy-Service Trading", The First IEEE International Conference on Multimedia Big Data (BigMM 2015), page 282-283, IEEE, 2015.
5. Ake Osothongs, Vorapong Suppakitpaisarn, and Noboru Sonehara. "A Proposed Method for Personal Attributes Disclosure Valuation: A Study on Personal Attributes Disclosure in Thailand", International Conference on Information Technology and Electrical Engineering (ICITEE 2015), page 408-413, IEEE, 2015.
6. Rohit Kumar Singh, Vorapong Suppakitpaisarn, and Ake Osothongs. "Improving Motivation in Survey Participation by Question Reordering", Pacific Rim Knowledge Acquisition Workshop (PKAW), page 231-240, Springer, 2016.

BIBLIOGRAPHY

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

[2] M. Otsuki and N. Sonehara, "Estimating the Value of Personal Information with SNS, Utility," in *Eighth International Conference on Availability, Reliability and Security (ARES)*, Regensburg, Germany, 2013.

[3] OECD, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.," 2013. [Online]. Available: https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

[4] Directive, E. U., 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 EC23,6, 1995.

[5] Office of the Australia Information Commissioner, "Privacy Act," [Online]. Available: https://www.oaic.gov.au/privacy-law/privacy-act/.

[6] Community Legal Information Centre(CLIC), "The meaning of "personal data" and the six data protection principles," [Online]. Available: http://www.clic.org.hk/en/topics/personalDataPrivacy/6_data_protection_principles/.

[7] The Office of the Privacy Commissioner of Canada, "A Guide for Individuals Protecting Your Privacy," [Online]. Available: https://www.priv.gc.ca/en/about-the-opc/publications/guide_ind/ .

[8] Civil Impulse, "S. 1332 — 109th Congress: Personal Data Privacy and Security Act of 2005.," 2014. [Online]. Available: http://www.govtrack.us/congress/bills/109/s1332.

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

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

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

[12] N. Arvind and S. Vitaly, “Myths and fallacies of personally identifiable information.,” *Communications of the ACM.,* pp. 53.6: 24-26., 2010.

[13] The World Economic Forum, “Unlocking the Value of Personal Data: From Collection to Usage,” 2013. [Online]. Available: https://www.weforum.org/reports/unlocking-value-personal-data-collection-usage.

[14] “Why do we need an EU data protection reform?,” 2013. [Online]. Available: http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/1_en.pdf.

[15] “Gartner IT Glossary,” [Online]. Available: http://www.gartner.com/it-glossary/big-data/.

[16] Statista Inc., [Online]. Available: https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/.

[17] J. Rose, O. Rehse and B. Röber, “The Vaue of our Digital Identity.,” 2012. [Online]. Available: http://www.lgi.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf.

[18] U.S. Attorney’s Office, “Miami Student Sentenced for Cyberstalking on Facebook and Instagram,” August 2016. [Online]. Available: 2016, https://www.justice.gov/usao-sdfl/pr/miami-student-sentenced-cyberstalking-facebook-and-instagram.

[19] K. Gammell, “My Facebook profile was stolen to get dates on Tinder - and there's nothing I can do,” May 2015. [Online]. Available: http://www.telegraph.co.uk/women/womens-life/11588667/Facebook-identity-theft-My-profile-was-stolen-to-get-dates-on-Tinder.html .

[20] G. G. Fuster and A. Scherrer, “Big Data and Smart Devices and their Impact on Privacy,” 21 09 2015. [Online]. Available: http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2015)536455.

[21] A. Nordrum, “The Internet of Fewer Things,” IEEE Spectrum, October 2016. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7572524 .

[22] A Federal Trade Commission, “Internet of Things: Privacy & Security in a Connected World,,” 2015 January. [Online]. Available: https://www.ftc.gov/

system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf .

[23] R. Montes, W. Sand-Zantman and T. Valletti, “The value of personal information in markets with endogenous privacy.,” 28 May 2015. [Online]. Available: https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2015/wp_tse_583_0.pdf.

[24] The World Economic Forum, “Personal Data: The Emergence of a New Asset Class,” February 2011. [Online]. Available: http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

[25] S. Kroft, “The data brokers: Selling your personal information,” 1 November 2014. [Online]. Available: http://www.cbsnews.com/news/the-data-brokers-selling-your-personal-information/.

[26] P. Boulin, “The Secretive World of Selling Data About You,” May 2016. [Online]. Available: http://europe.newsweek.com/secretive-world-selling-data-about-you-464789.

[27] S. E. Gindin, “Nobody Reads Your Privacy Policy or Online Contract? Lessons Learned and Questions Raised by the FTC's Action Against Sears,” *The Northwestern Journal of Technology and Intellectual Property ,* vol. 8, no. 1, 2009.

[28] J. Gomez, T. Pinnick and A. Soltani, “KNOWPRIVACY,” June 2009. [Online]. Available: http://knowprivacy.org/report/KnowPrivacy_Final_Report.pdf .

[29] CBS News, “Beware downloading some apps or risk “being spied on”,” February 2016. [Online]. Available: http://www.cbsnews.com/news/mobile-phone-apps-malware-risks-how-to-prevent-hacking-breach/.

[30] N. Mojtahedi, “Popular free apps compromise your personal information and security,” December 2015. [Online]. Available: http://globalnews.ca/news/2391430/popular-free-apps-compromise-your-personal-information-and-security/.

[31] Federal Trade Commission, “VIZIO to Pay $2.2 Million to FTC, State of New Jersey to Settle Charges It Collected Viewing Histories on 11 Million Smart Televisions without Users’ Consent,” February 2017. [Online]. Available: https://www.ftc.gov/news-events/press-releases/2017/02/vizio-pay-22-million-ftc-state-new-jersey-settle-charges-it.

[32] M. Gijzemijter, "'Facebook is gathering personal information without consent,' says Belgian privacy watchdog," May 2015. [Online]. Available: http://www.zdnet.com/article/facebook-is-gathering-personal-information-without-consent-belgian-privacy-watchdog/ .

[33] K. J. O'Brien, "Germany Fines Google Over Data Collection," April 2013. [Online]. Available: http://www.nytimes.com/2013/04/23/technology/germany-fines-google-over-data-collection.html.

[34] The Nordic Page, "Google Fined for Illegal Collection of Personal Information in Norway," August 2012. [Online]. Available: http://www.tnp.no/norway/economy/3138-google-fined-for-illegal-collection-of-personal-information-in-norway.

[35] C. Savage and J. Weisman, "N.S.A. Collection of Bulk Call Data Is Ruled Illegal," May 2015. [Online]. Available: https://www.nytimes.com/2015/05/08/us/nsa-phone-records-collection-ruled-illegal-by-appeals-court.html.

[36] A. Travis, "UK security agencies unlawfully collected data for 17 years, court rules.," October 2016. [Online]. Available: https://www.theguardian.com/world/2016/oct/17/uk-security-agencies-unlawfully-collected-data-for-decade.

[37] Office of the Australian Information Commissioner(, "What does 'trading in personal information' mean?," 1 November 2014. [Online]. Available: http://opc.joomla-prime.icemedia.com.au/index.php?option=com_content&view=article&id=721&Itemid=1570.

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

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

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

[41] J. Angwin, "Privacy Tools: Opting Out from Data Brokers," January 2014. [Online]. Available: https://www.propublica.org/article/privacy-tools-opting-out-from-data-brokers.

[42] P. Dixon and R. Gellman, “Consumer Tips: World Privacy Forum’s Top Ten Opt Outs,” June 2016. [Online]. Available: https://www.worldprivacyforum.org/2015/08/consumer-tips-top-ten-opt-outs/ .

[43] P. Dixon, “Congressional Testimony: What Information Do Data Brokers Have on Consumers?,” 2013 December. [Online]. Available: https://www.worldprivacyforum.org/2013/12/testimony-what-information-do-data-brokers-have-on-consumers/ .

[44] A. Kearney, “Rethinking Personal Data: A New Lens for Strengthening Trust,” 2014 November 2014. [Online]. Available: http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf.

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

[46] A. Felt and D. Evans, “Privacy protection for social networking platform,” in *Workshop on Web 2.0 Security and Privacy (W2SP’08)*, 2008.

[47] M. Chahal, “Consumers are ‘dirtying’ databases with false details,” July 2015. [Online]. Available: https://www.marketingweek.com/2015/07/08/consumers-are-dirtying-databases-with-false-details/.

[48] E. Aïmeur and D. Schonfeld, “The ultimate invasion of privacy:Identity theft,” in *2011 Ninth Annual International Conference on Privacy, Security and Trust (PST)*, 2011.

[49] K. Heather, “83 million Facebook accounts are fakes and dupes, Available,” August 2012. [Online]. Available: http://edition.cnn.com/2012/08/02/tech/social-media/facebook-fake-accounts/.

[50] B. Wheeler, “Give social networks fake details, advises Whitehall web security official,” October 2012. [Online]. Available: http://www.bbc.co.uk/news/uk-politics-20082493.

[51] C. Gustke, “Which countries are better at protecting privacy?,” June 2013. [Online]. Available: http://www.bbc.com/capital/story/20130625-your-private-data-is-showing.

[52] A. Marwick, “How Your Data Are Being Deeply Mined,” pp. 22-25, 14 January 2014.

[53] D. Zax, “Is Personal Data the New Currency?,” MIT Technology Review, 2011 November. [Online]. Available: https://www.technologyreview.com/s/426235/is-personal-data-the-new-currency/.

[54] A. Krotoski, “Battle for the internet Big Data age puts privacy in question as information becomes currency,” April 2012. [Online]. Available: https://www.theguardian.com/technology/2012/apr/22/big-data-privacy-information-currency .

[55] A. Musayeva, “Is data the new currency?,” July 2015. [Online]. Available: http://tedx.amsterdam/2015/07/is-data-a-new-currency/.

[56] T. Cochran, “Personal Information Is the Currency of the 21st Century,” May 2013. [Online]. Available: http://allthingsd.com/20130507/personal-information-is-the-currency-of-the-21st-century/ .

[57] N. Chandrasekaran, “Is data the new currency?,” August 2015. [Online]. Available: https://www.weforum.org/agenda/2015/08/is-data-the-new-currency/.

[58] Future Foundation, “Data privacy: what the consumer really thinks,” December 2015. [Online]. Available: http://dma.org.uk/uploads/ckeditor/Data-privacy-2015-what-consumers-really-thinks_final.pdf.

[59] E. Steel, C. Locke, E. Cadman and B. Freese, “How much is your personal data worth?,” 1 November 2014. [Online]. Available: http://www.ft.com/intl/cms/s/2/927ca86e-d29b-11e2-88ed-00144feab7de.html.

[60] K. Burney, J. Brehm and K. Robinson., “Valuing Identity in Today’s Digital World: The business case for defining digital identity and how to value it correctly.,” 2013. [Online]. Available: https://www.unboundid.com/company/news/press/-2013/20130730.php.

[61] H. McCracken, “Cloudsweeper’s Gmail Security Audit Is Alarming and Useful.,” 1 November 2014. [Online]. Available: http://techland.time.com/2013/06/27/gmail-security/.

[62] Trend Micro, “How Much is Your Personal Data Worth? Survey Says...,” April 2015. [Online]. Available: https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/how-much-is-your-personal-data-worth-survey-says .

[63] J. Staiano, N. Oliver, B. Lepri, R. Oliveira, M. Caraviello and N. Sebe, "Money Walks: A Human-Centric Study on the Economics of Personal Mobile Data," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing 2014*, 2014.

[64] S. Joshana and X. B. Yan., "Investigating Effects of Monetary Reward on Information Disclosure by Online Social Networks Users," in *47th Hawaii International Conference on System Sciences (2014)*, Hawaii, 2014.

[65] 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.,* vol. 29, pp. 350-353, 2002.

[66] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression.," *International Journal of Uncertainty. Fuzziness and Knowledge-Based Systems.,* pp. 10.05: 571-588., 2002.

[67] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity.," *ACM Transactions on Knowledge Discovery from Data.,* vol. 1, no. 1, p. 3, 2007.

[68] N. Li, T. Li and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity.," *ICDE,* vol. Vol. 7, 2007.

[69] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas and A. Zhu, "Approximation algorithms for k-anonymity," *Journal of Privacy Technology,* 2005.

[70] K. M. Choromanski, T. Jebara and K. Tang, "Adaptive Anonymity via b-Matching.," *Advances in Neural Information Processing Systems.,* 2013.

[71] M. Geist, "The Expansion of Personal Information Disclosure Without Consent: Unpacking the Government's Weak Response to Digital Privacy Act Concerns," October 2014. [Online]. Available: http://www.michaelgeist.ca/2014/10/expansion-personal-information-disclosure-without-consent-unpacking-governments-weak-response-digital-privacy-act-concerns/ .

[72] The Office of the Privacy Commissioner of Canada, "A Guide for Individuals Protecting Your Privacy," [Online]. Available: https://www.priv.gc.ca/en/about-the-opc/publications/guide_ind/.

[73] BC Freedom of Information and Privacy Association, “When Consent Isn't Needed,” [Online]. Available: http://www.healthinfoprivacybc.ca/consent/when-consent-isnt-needed.

[74] Office of the Official Information Commission, “Disclosure of personal information without the consent of the data owner,” [Online]. Available: http:// www.oic.go.th/web2014/en/disclosure_pi_without_consent_data_subject.htm.

[75] R. Gellman, “Privacy in the clouds: risks to privacy and confidentiality from cloud computing.,” February 2009. [Online]. Available: http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf.

[76] K. Gibson, “Large scale emergencies and personal information – can the Privacy Act cope?,” [Online]. Available: https://www.privacy.org.nz/assets/Files/Codes-of-Practice-materials/Katherine-Gibson-paper-on-Earthquake-code-10-05-11.pdf.

[77] Office of the Information Commissioner, “Privacy flexibilities in the management of disaster,” [Online]. Available: https://www.igem.qld.gov.au/innovation/ Documents/Privacy-flexibilities-in-the-management-of-disaster-events.pdf.

[78] S. Aguinaga and C. Poellabauer, “Method for privacy-protecting display and exchange of emergency information on Mobile devices,” in *2012 International Conference on Collaboration Technologies and Systems (CTS)*, 2012.

[79] B. Woods, “thenextweb.com,” 2013. [Online]. Available: http://thenextweb.com/ insider/2013/09/17/whats-the-true-value-of-your-personal-data-meet-the-people-who-want-to-help-you-sell-it/.

[80] J. Cook, “Startup Spotlight: Enliken helps consumers control personal data and market it to advertisers,” March 2013. [Online]. Available: http://www.geekwire.com/2013/enliken/.

[81] Handshake, “Handshake,” 2013. [Online]. Available: http://www.handshake.uk.com/.

[82] A. Osothongs and N. Sonehara, “A proposal of personal information trading platform (PIT): A fair trading between personal information and incentives,” in *Conference on Digital Information and Communication Technology and its Applications(DICTAP)*, 2014.

[83] W3C, “Platform for Privacy Preferences (P3P) Project,” 2007. [Online]. Available: http://www.w3.org/P3P/.

[84] Electronic Privacy Information Center, “Pretty Poor Privacy: An Assessment of P3P and Internet Privacy,” 2000. [Online]. Available: https://epic.org/reports/prettypoorprivacy.html.

[85] K. El-Khatib, “A Privacy Negotiation Protocol for Web Services,” in *Proceedings of the International Workshop on Collaboration Agents: Autonomous Agents for Collaborative Environments (COLA)*, 2003.

[86] D. Walker, E. Mercer and K. Seamons, “Or Best Offer: A Privacy Policy Negotiation Protocol,” in *2008 IEEE Workshop on Policies for Distributed Systems and Networks*, 2008 .

[87] A. Yassine and S. Shirmohammadi, “Privacy and the Market for Private Data A Negotiation Model to Capitalize on Private Data,” in *AICCSA*, Qatar, 2008.

[88] A. Ukil, S. Bandyopadhyay, J. Joseph, V. Banahatti and S. Lodha, “Negotiation-based Privacy Preservation Scheme in Internet of Things Platform,” in *ACM SECURIT*, 2012.

[89] O. Kwon, “A pervasive P3P-based negotiation mechanism for privacy-aware pervasive e-commerce,” *Decision Support Systems,* vol. 50, no. 1, pp. 213-221, December 2010.

[90] Facebook, “Updates to Facebook Login,” 1 December 2013. [Online]. Available: http://newsroom.fb.com/News/696/Updates-to-Facebook-Login .

[91] Facebook, “Graph API Version 2.8,” 1 December 2014. [Online]. Available: https://developers.facebook.com/docs/graph-api/reference/user.

[92] Gephi NGO, “Gephi – the open graph viz platform,” 6 November 2014. [Online]. Available: http://gephi.org.

[93] V. D. Blondel, J. L. Guillaume, R. Lambiotte and L. Etienne, “Blondel, V. D., Guillaume, J. L., Lambiotte, R. and Etienne,L. 2008. Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment,* p. 10: P1000, 2008.

[94] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, Introduction to algorithms, MIT Press and McGraw-Hill, 2009, p. 655–657.

[95] B. L. Welch, The generalization of Student's problem when several different population variances are involved., Biometrika, 1947, pp. 34.1:28-35.

[96] M. Shacklett, "Big data wake-up call: Increased online privacy concerns require risk management.," 1 November 2014. [Online]. Available: http://www.techrepublic.com/article/big-data-wake-up-call-increased-online-privacy-concerns-require-risk-management/.

[97] R. Kumar Singh, V. Suppakitpaisarn and A. Osothongs, "Improving Motivation in Survey Participation by Question Reordering," in *Pacific Rim Knowledge Acquisition Workshop (PKAW)*, Phuket, 2016.

[98] C. Suvarnapunya and P. Jarupunphol, "Data protection in Thailand: overview," August 2016. [Online]. Available: http://us.practicallaw.com/0-520-0782.

[99] W. Zeldin, "Thailand: Digital Ministry Established as Part of National Digital Economy Plan," June 2016. [Online]. Available: Available: http://www.loc.gov/law/foreign-news/article/thailand-digital-ministry-established-as-part-of-national-digital-economy-plan/ .

[100] D. Piper, "Compare data protection laws around the world," January 2017. [Online]. Available: https://www.dlapiperdataprotection.com/index.html ?t=law&c=TH .

[101] "Thailand: Draft of Personal Information Protection Act," [Online]. Available: https://ictlawcenter.etda.or.th/de_laws/detail/de-laws-data-privacy-act.

[102] Google, "Google Forms," [Online]. Available: https://www.google.com/intl/en/forms/about/.

ABOUT AUTHOR

Name	Ake Osothongs
Birthplace	Bangkok
Nationality	Thai
Educations	2009/7 Master of Science in Information Technology Asian Institute of Technology (AIT), Thailand 2003/4 Bachelor of Science in Computer Science Bangkok University, Thailand
Experiences	2011/10 – Present: Officer, Defence Information and Space Technology Department, Office of the Permanent Secretary for Defence, Ministry of Defence, Thailand 2010/7-2011/7: Programmer, DST Worldwide Service, Thailand 2010/1-2010/4: Technical consultant, ADP Dealer service (Thailand) Ltd., Thailand 2005/4 – 2007/6: System Analyst, ExxonMobil Ltd., Thailand