Electronic Preprint for Journal of Information Processing Vol.25

Regular Paper

Privacy Disclosure Adaptation for Trading between Personal Attributes and Incentives

Ake Osothongs[1,a)] Vorapong Suppakitpaisarn[2,b)] Noboru Sonehara[3]

Received: March 10, 2016, Accepted: September 6, 2016

Abstract: Products and services nowadays need personal information from consumers in order to personalize their goods to best fit consumers. At the present, the online environment is the biggest source of consumers' personal information. However, online privacy has become the major concern of consumers. A personal information trading platform has been proposed as a medium for collecting consumers' personal information in exchange for monetary incentive. This study proposes a new approach to requesting personal attributes which can adapt with consumers' personal information disclosure behavior and aims to increase the disclosure of personal information without increasing of monetary incentive. To develop this new adaption method, we developed the valuation of a personal information method without using currency. The probability and graph mining techniques were used to valuating personal attributes. Then, we displayed the relationships of personal attributes disclosure in the hierarchy and proposed a method for valuating personal information disclosure. The valuation method was used in the evaluations, which were compared with the disclosure of personal information results from the consumers. After the evaluation was completed, the result showed that the new approach can significantly increase the disclosure of consumers' personal information.

Keywords: privacy, privacy disclosure, privacy trading

1. Introduction

Every time people log on to the Internet they leave footprints. Their activities, as well as their personal and individual data, are collected and stored by the service providers. This data plays an important role in this digital age because it usually consists of personal information. Personal information becomes a data source for better understanding customer profiles, thus improving customer satisfaction for services and products received. Nowadays, many service providers rely on this type of data. For example, online service providers such as e-commerce use personal information to predict consumer demand for their product stocks, while Internet service providers use the data to manage and improve the performance of their networks. Many techniques and models have been proposed to manage and analyse this data. Service providers rely on personal information which plays an important role in online technology.

Privacy and security problems recently have been discussed because of the rise of the age of the Internet of Things. Consumers concern has risen dramatically with increasing awareness regarding privacy risks [1]. In 2015, a survey by the National Cyber Security Alliance (NCSA) and the data privacy company TRUSTe, showed that over 90 percent of consumers are aware and worried about their online privacy. They are especially concerned when companies collect their personal information and share it with third parties [2]. Previous works found that consumers resist cooperation with organizations that they do not trust; they may decline to disclose personal information if they feel uncomfortable [3]. Personal information disclosure is more complicated than being simply the decision of the consumer as to whether to provide it or not. The privacy issue always develops into a paradox, because even if consumers are aware and concerned about their privacy, they still casually disclose their personal details in the online environment.

Consumers usually trade-off their personal information against the convenience on the online environment. For example, they provide many personal information attributes when using Social Network Service (SNS) accounts for logging into third party websites to reduce registration and log in time. Furthermore, consumers also allow service providers to collect information when installing new mobile phone applications. Thus, consumers usually agree to disclose their personal information as a trade for other incentives that they desire. In other words, consumers generally release their personal information when they really need to use online services or products [4].

Many education providers, telecommunication service providers and Internet service providers collect personal information through consumer site usage. However, they require legal consent from the consumer to use this data. Service providers often use monetary incentives to attract consumer attention to endorse the disclosure of personal information. Trading platforms have been proposed as mediums between personal information and incentive [5]. The service provider prepares an incentive, requesting the consumer to disclose a particular personal attribute. The consumer receives the offer and then

1 SOKENDAI (The Graduate University of Advanced Studies), Hayama, Kanagawa 240–0193, Japan
2 The University of Tokyo, Bunkyo, Tokyo 113–0033, Japan
3 National Institute of Informatics, Chiyoda, Tokyo 101–8430, Japan
a) ake.osothongs@gmail.com
b) vorapong@is.s.u-tokyo.ac.jp

makes a decision. He/she will receive the incentive on agreement to disclose the personal attribute.

1.1 Our Contribution

We observe that the Internet service providers sometimes ask consumers to provide attributes that most consumers do not want to disclose such as national ID. In that case, the trading activities are not effective as the service providers will obtain a minimal amount of information, even if their services are useful.

To reduce that ineffectiveness, in Section 3, we conduct a survey on 532 participants, evaluate each personal attribute, and develop software to inform the providers of the importance of the attributes in consumers' point of view.

Discussed in Section 5, there are many works proposing valuations on personal attributes. Most of the works aim to find the monetary values of the attributes. However, the values are very sensitive to context. The values depend on the cost of living of the period when the survey is conducted, and are also affected by the survey participants' average cost of living.

Instead of the monetary values, we construct a graph structure based on our survey results. Each node in the graph represents a personal attribute. Then, we calculate the relative importance for each pair of personal attribute using the conditional probability, and draw an edge from a relatively more important node to a relatively less important node. By that, we obtain a graph that shows the relationship between each personal attribute. The graph is insensitive to context, as we have a similar one when we construct a graph from the survey results obtained from only male participants or only female participants.

We develop the software to visualize the graph. To verify that the graph and the software is informative, we ask 14 service providers to use our software. Then, we ask them to give us a rating based on how much information on the consumers' personal attributes evaluation they got from the software. The average rating we have from the survey is 4.0 out of 5.

While we aim to promote the trading of information by increasing the service providers' awareness on the importance of personal attributes in Section 3, in Section 4, we aim to promote the trading by a question order that can increase the consumers' participation in the trading. To obtain the order, we perform the topological sorting on the graph obtained from Section 3, and perform the preorder traversal on the tree obtained from the sorting.

We conduct an experiment on 160 consumers to show that our question order can increase the consumers' participation. Compared to other question order, the survey results indicate that our order can increase the information provided from consumers by 18.25% with the p-value equal to 0.0591.

The question order can be improved based on the demographical data of the consumers. Instead of using the static data obtained from 532 participants in Section 3, we construct a question order using only a subset of participants which have the same demographical data with the consumers. Our experiment shows that, by this adaptive approach, the improvement is increased from 18.25% to 30.11% and the p-value is decreased.

Parts of this work have been published in ACM IMCOM-ICUIMC2015 [6], IEEE BigMM2015 [7], and ICITEE2015 [8].

2. Definition

The term "***personal information***" is used in this study. The definition of the term has been discussed extensively, but the precise definition remains unclear and varies based on what may encompass the term [9], [10]. There are many similar terms such as personal information, personal data, and personal identifiable information, which are used widely in the same context. Further, the definitions of such terms have changed over time with the development of technology. In the 1900s to early 21st century, most computer systems were standalone systems and most databases were still offline. The definition of personal information usually referred to the information that could directly or indirectly identify an individual person [11]. However, changes in technology have affected changes in the definition of the term. Nowadays, most computer systems connect to the Internet and database systems are not singularly located. Service providers may collect many pieces of personal information and combine them to identify an individual. Researchers have already found that it is possible to re-identify a particular person from just small pieces of seemingly unimportant personal information [12], [13].

Therefore, "***personal information***" is defined as any information that can directly or indirectly relate to a specific person, regardless of the sources. This study also used the term "personal attribute" when describing specified personal information.

In this study, we focused on the ***trading activities***. In these activities, we aim to place a ***trading platform*** between two actors, the service provider and the consumer. A ***service provider*** is any public or private organization that provides online services to its consumers. They collect personal information from their services to their databases, but still cannot use it without the consent of consumers. A ***consumer*** is any person who uses online services from a service provider.

3. Robust Valuation Method for Personal Attributes

Section 3.1 describes a method using conditional probabilities to construct a graph that identifies the relative importance between personal attributes. Section 3.2 discusses the validity and interpretation of this graph and Section 3.3 describes the software used for visualizing the graph, and its value to service providers.

3.1 Methods

One of the problems of the trading platform is the estimation of personal information value. Previous studies showed that qualitative value is sensitive to context [14], [15]. The facts tend to show that consumers disclose personal information more freely when they feel comfortable and happy to accept incentives from service providers, which may not be purely financial incentive. We propose a valuation method for personal information using a graph-mining technique. The proposed method aimed for measurement and comparison of personal attributes from the consumer's point of view without using a quantitative value. Instead, the method used a probability technique when consumers disclosed personal information and a graph-mining technique as follows:

1) A dataset was collected from 532 Thai consumers in Ref. [5].

Table 1 Demographic details of subjects who participated in the survey.

Attribute	Value	Percentage
Gender	Male	48.4%
	Female	52.1%
Education	High School	5.4%
	College	2.8%
	Bachelor's Degree	47.6%
	Graduate	40.5%
Age	15-20	3.7%
	21-30	39.3%
	31-40	47.8%
	41-60	8.6%
	Over 60	0.03%
Occupation	Student	16.6%
	Self-employed	14.0%
	Private company	32.0%
	Government Officer	27.2%
	State Enterprise	2.8%
	Unemployed/Housemaid	3.5%
	Other	3.5%

The questionnaire survey consisted of two parts, which are demographic information and Likert scale five-level questions regarding comfort levels when disclosing the 33 distinct personal attributes shown in Table 2. The demography of the survey participants known from the first part of the survey is shown in **Table 1**. The results from the second part of the survey were separated into two conditions. If the answers to the questions were 1, 2 or 3 the condition was termed *disclose*. If the answers were 4 or 5 the condition was termed *protect*.

2) Calculate the probability for protection of a personal attribute given by other attributes. We adopted Bayes' formula for calculating the conditional probabilities. Let A and B be personal attributes. The possibility that a consumer discloses personal attribute B when he/she discloses personal attribute A as $P(B\,|\,A)$ can be calculated as follows:

$$P(B\,|\,A) = \frac{P(A \bigcap B)}{P(A)}$$

where $P(A \bigcap B)$ is the probability that a consumer protects both personal attributes A and B and $P(A)$ is the probability that the consumer protects personal attribute A.

For example, from the total of 532 subjects 337 chose not to disclose their name. Among those 337 subjects, 324 chose not to disclose their home number. Therefore,

$$P(Name) = \frac{337}{532} \approx 0.633$$

and $P(Name \bigcap Homenum) \approx 0.609$. We have

$$P(Homenum\,|\,Name) \approx \frac{0.609}{0.633} = 0.961,$$

which means that 96.1% of subjects who did not disclose their name also chose to protect their national IDs. Similarly,

$$P(Officenum\,|\,Nickname) = \frac{104}{113} \approx 0.920,$$

because 113 subjects chose not to disclose their nicknames and 104 of these chose not to disclose their office number.

3) A probability $P(B\,|\,A)$ close to one indicated that most survey participants who chose to protect A also chose to protect B. Thus, this implied that personal attribute B was more important than personal attribute A.

A directed graph was constructed where each node represented a personal attribute. An edge was drawn from a node representing B to a node representing A when $P(B\,|\,A) > 0.95$, i.e., an edge from B to A existed in the graph if and only if B was more important than A. The graph is shown in **Fig. 1**.

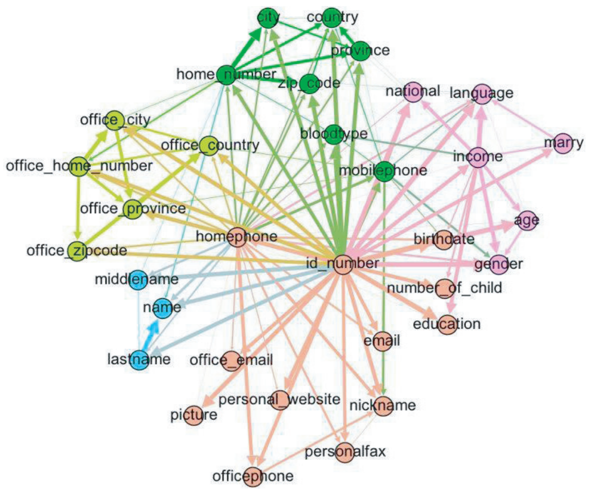

Fig. 1 The graph obtained from step 3 of the method.

3.2 Interpretation of Graph Construction

The graph in Fig. 1 shows that most users considered national ID, home phone number and mobile phone number more important than the remaining attributes, as there are edges from these attributes to almost all the others. Attributes such as age, gender, language or nickname were considered less important as all edges corresponding to these nodes came from other nodes. This conformed to the general facts.

The community structure of the graph was then studied using a clustering algorithm that maximized the graph modularity proposed in Ref. [16]. The algorithm divided the graph into five clusters shown as different colours in Fig. 1. The modularity was obtained using the method from the graph at 0.26, which was significantly high considering that the graph in Fig. 1 was very sparse. This result indicated that the graph had a community structure.

To test the hypothesis that each cluster contained attributes which were semantically similar, 33 personal information attributes were manually classified into five categories. Each category contained attributes with similar meaning listed as follows:

1) **Name**: Information related to personal name
2) **Home address**: Information related to home address
3) **Office address**: Information related to office address
4) **Contact information and personal identifiable information**: Other information that can identify a single person
5) Other personal information such as demographical in-

Table 2 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 number, City, Province, Country, Zip code	Office 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, Nationality, Age, Income, Language, Marital status, Education, Number of children, Blood type	Gender, Nationality, Age, Income, Language, Marital status	1	0.67

formation

Contact information and personal identifiable information were combined because both types can identify an individual.

The classification was compared with the clustering result in **Table 2**. The results were very similar and 84.84% of the personal information was correctly classified. The precision and recall were calculated for each category of personal information. Results also showed the restricted relationships among the personal information attributes in every category and an association in user decisions to protect two attributes in the same semantic group.

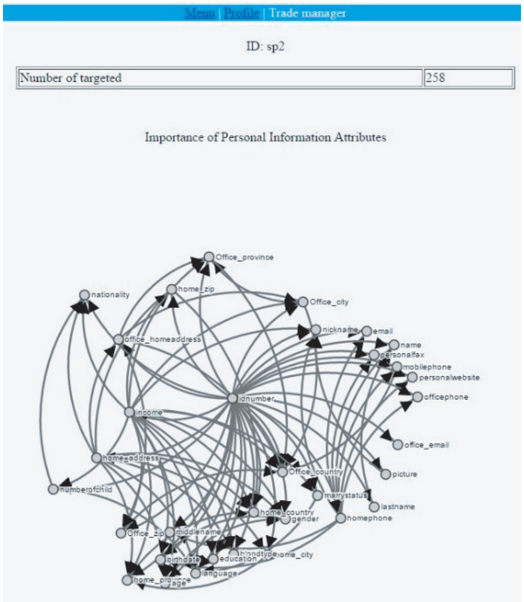

Fig. 2 Screenshot of the graph visualization software.

3.3 Graph Visualization

However, the graph is complicated and service providers might find it difficult to read and interpret. Therefore, we develop the software to visualize the graph. This software showed only a subgraph containing personal attributes that related to the personal attribute where the users' mouse was on. A screenshot of the software is shown in **Fig. 2**.

We also provide the service providers an option to view the graph constructed from a specific group of consumers. For example, the service providers can choose to view a graph constructed from 258 female participants.

A sample group of 14 e-commerce website owners was invited to evaluate the system. Firstly, they were asked to create five campaigns to trade between personal information and incentives without using the new software. Each selected five personal attributes per campaign with different groups of target markets. Secondly, they performed the same task with support from the software. Once both tasks were completed they were asked to complete a questionnaire survey showing satisfaction rated on a scale of 1 to 5.

The results showed that the value of personal attributes affected the service providers' decisions to request personal information. The service providers changed their decisions when they understood the value of information to consumers after using the software. Service providers avoided requesting personal attributes that were of high importance to the consumers. Statistically, the participants were satisfied with the software, the average satisfaction rate was 4.0 and the standard deviation was 0.7.

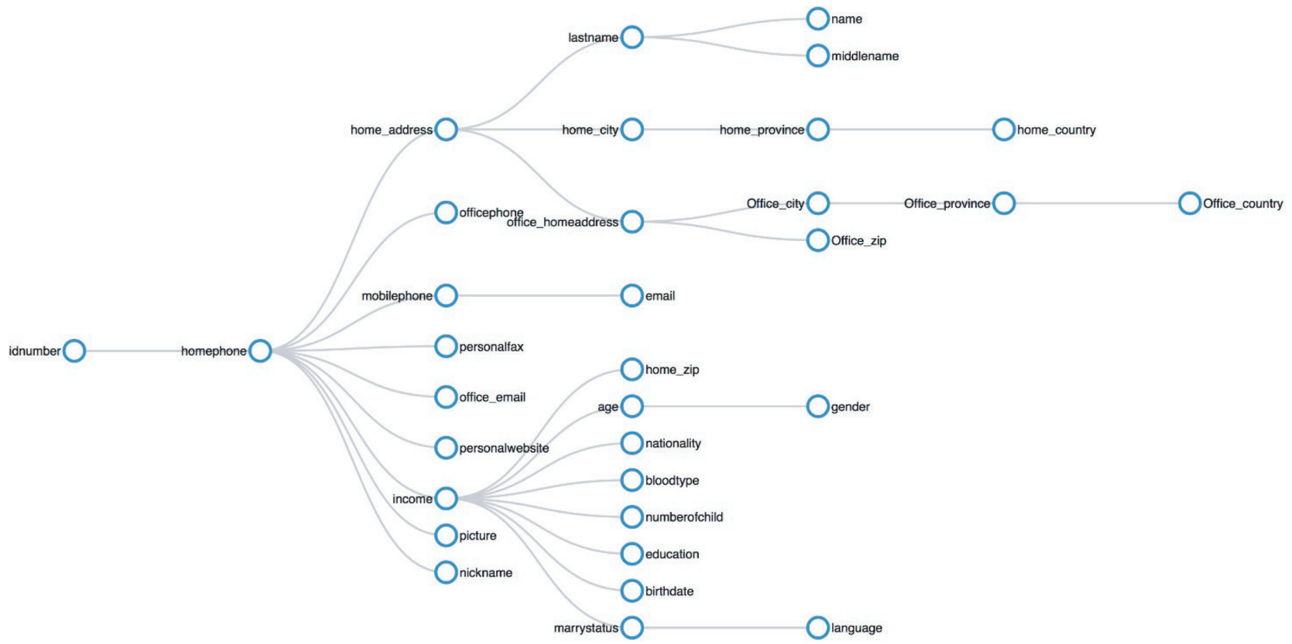

Fig. 3 The results tree graph.

4. Conversion to the Value of Disclose

Although the graph obtained in the previous section was robust against the evaluation context, it was relatively hard to use in several applications. Therefore, a technique was proposed to transform the graph into numerical values called the *Value of Disclosure* (*VD*). Section 4.1 describes the method used to calculate *VD*. Section 4.2 describes software that utilizes the value and Section 4.3 evaluates the value and the developed software.

4.1 Methods

1) The graphs obtained in the previous section were converted to trees using topological sorting (c.f., Ref. [17]). This obtained a tree with the root nodes as the most important personal attributes and leaves to represent less important personal attributes. A node with a large outdegree in the graph was an important node, so the topological sort algorithm started at the edge with the largest outdegree. Nodes with large indegrees in the graph were usually not important nodes. Since they were visited after all nodes which have edges to them, they tend to be leaves that are far from the root node.

   By applying a topological sort algorithm, the graph in Fig. 2 was transformed into the tree shown in **Fig. 3**.

2) Let A be a node in the graph and a personal attribute and let R be a root of the tree obtained in 1). If $A = R$, then

$$VD_A := 1.$$

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

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

   where $W_i = P(A_i \,|\, A_{i-1}) = \frac{P(A_i \cap A_{i-1})}{P(A_{i-1})}$.
   By the argument in Section 3.1, we know that W_i represents the relative importance of A_i compared to A_{i-1}. The value of VD_A is then a value representing the relative importance of the attribute A compared to the most important attribute R. We do not directly assign $P(A \cap R)/P(R)$ to VD_A, because A and R do not have a strong relationship when there is no edge between A and R and we strongly believe that the relative importance W_i should be calculated only from two attributes with a strong relationship. The results obtained from this calculation are shown in **Table 3**.

4.2 Experiment

To verify the method in Section 4.1 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 also Thai Internet users. The participants were invited to register themselves 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 that each could 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 found a finish page. There were 33 personal attributes questioned in this application, which was the same number of personal attributes set for the calculated *VD*. **Figure 4** shows a screen shot of the web application when asking the questions to participants.

The order of personal attributes questioned in this research aimed to improve trading between personal information and monetary incentives. This study separated the participants into two

Table 3 Value of personal attribute disclosure.

Rank	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

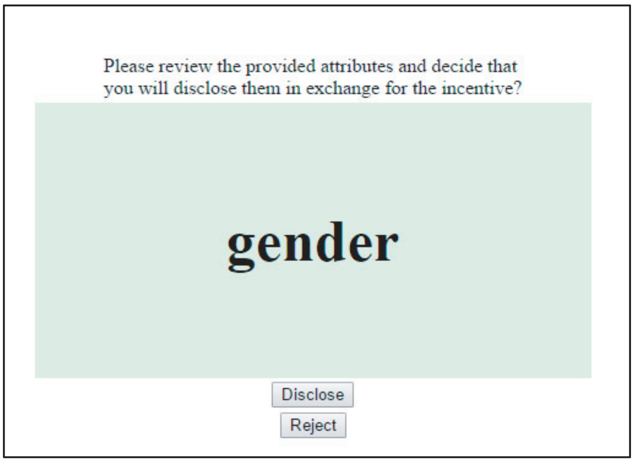

Fig. 4 Screenshot of the web application asking a disclosure question.

groups. Each group used the web application that questioned a different set of personal attributes. We called the two sets of questions "***top-down approach***" and "***bottom up approach***." Our

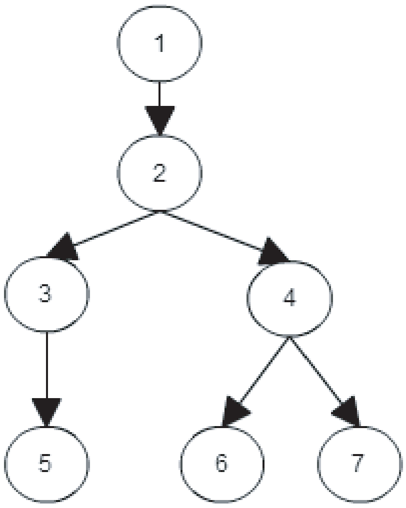

Fig. 5 Example of a tree for the ordering approach.

top-down approach *used the pre-order traversal* (c.f., Ref. [17]) 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* (c.f., Ref. [17]) to order the personal attribute questions from the leaf 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, **Fig. 5** is a tree containing 7 personal attributes, represented by node 1 to node 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 started 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.

4.3 Evaluation and Results

From previous studies, we believed that service providers can gain more benefits when they understand the consumer's disclosure of personal information. In this study, we conducted two consecutiveexperiments. In the first experiment, we invited 100 participants to use our web application. The invited group are internet users in Thailand because we had collected, used, constructed and calculated the tree and *VD* from this group of users in Thailand. A different group of users may affect their 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 participants were separated into two groups. Each group completed a different approach from top-down and bottom-up approaches.

The experiments were completed by participants and their answers collected. We compared the results with the total *VD* when all personal attributes are disclosed as 18.735. The average of *VD* from each group of participants is shown in **Table 4**. When the top-down approach was used, the average of total *VD* is 11.263, which is 60.12 percent. When the bottom-up approach was used, the average of total *VD* is 9.5254, which is 50.84%.

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 [18]. The *p*-value obtained from the calculation was 0.0591. Although the value was still higher than a conventional criteria at 0.05, we believe that the value was small enough to conclude that the top-down approach was better than the bottom-up approach.

We conducted the second experiment in this research. The result from the top-down approach was used as a baseline in this second experiment that aimed to improve the consumers' disclosure of personal information. We invited more participants for this experiment and we adapted the order of personal attributes in our web application by their profile. In this study, we selected their gender as a criterion because we found the difference on their disclosure. The new trees were constructed for male and female, and then the ordering of personal attributes was 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, if the consumer participating in the survey was female an order was constructed from the 258 female participants in Section 2.1. The results obtained following the improvement are shown in **Table 5**. We will call this technique ***the adaptive approach***.

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 [18]. Unfortunately, the *p*-value obtained from the calculation was 0.180. Although the top-down adaptive approach had a significantly higher average disclosure, we 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 percentages of personal attribute disclosure of participants were calculated and shown in **Fig. 6**. The graph displayed the difference in results of consumer disclosure of each personal attribute between the top-down approach, bottom-up approach and the enhanced approach. The results of the top-down approach showed that participants disclose their personal attributes easily when the web application started the question from high *VD* attributes to low *VD* attributes. Participants may disclose the 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 decreases

Table 4 Result of top-down and bottom-up approaches.

Approach	Average	SD	#Participants
Top-down	11.263	4.533	49
Bottom-up	9.525	4.563	51

Table 5 Result of New Approach.

Approach	Average	SD	#Participants
Top-down with the adaptive approach	12.393	4.192	60

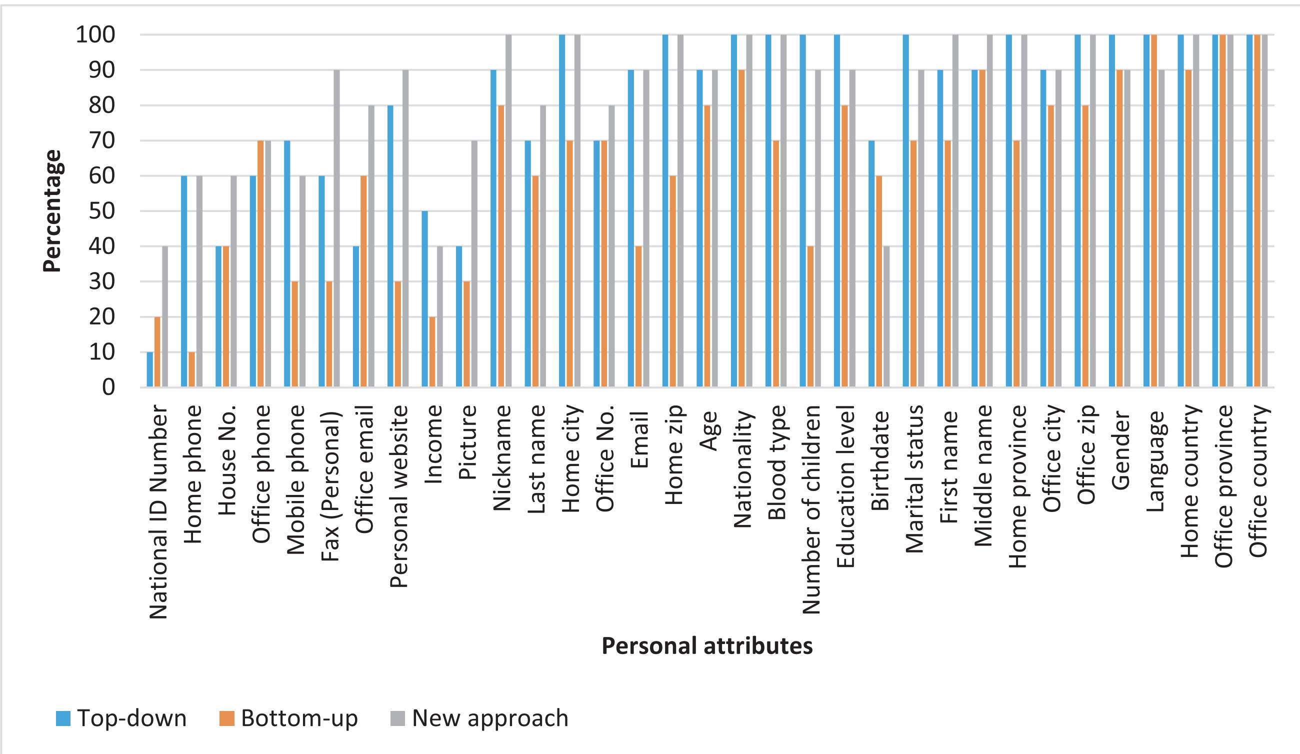

Fig. 6 Experiment result and comparison.

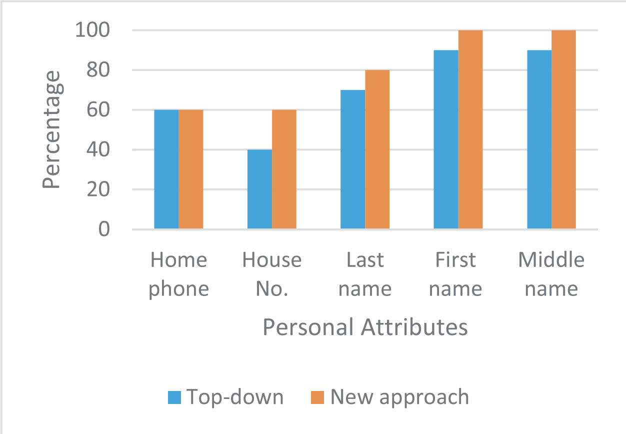

Fig. 7 Example of the comparison result.

in the bottom-up approach.

Our adaptive approach has been used to improve the top-down approach. The results of the new approach showed that the disclosure of personal attributes of participants can be increased. From the graph, the disclosure of personal attributes of 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 new approach come out and support our assumptions that consumers disclose their personal attributes in a hierarchical form and the personal attributes have semantic similarity among them. **Figure 7** is an example of the disclosure results in percentage form from a set of personal attributes. These personal attributes have semantic similarity and hierarchy in our graph in Fig. 3. From our calculation, home phone has the highest position in the tree and participants disclosed the personal attributes in the form of a hierarchy under the home phone node until reaching the leaf node, which comprises first name and middle name. The results in Fig. 7 show that the disclosure ratio was increasing when using the adaptive approach.

5. Related Work

Many researchers have focused on the problem that is the valuation of personal information. Every day, data brokers collect, sell or transfer personal information to third parties. This fact shows that information can be treated the same as a commodity. However, the value of personal information is difficult to calculate. Several studies suggested estimation methods for the valuation of personal information in a financial context (money). The results of these studies vary based on the proposed estimation methods. The value of personal information can be very high in one study while very low in another. The actual value of personal information is still difficult to estimate because people do not disclose their information only for tangible incentives; they also disclose their personal information for intangible incentives too.

A study from the Financial Times estimated personal information's worth for each person using pricing data from the industry in the US. The results showed that personal information for the average person was worth less than one dollar in the US. The value of consumer's personal information increases when a person reaches a turning point in their life. For example, they need to find something new and demand for it [19]. Moreover, data brokers usually sell personal information such as the email addresses and contact information of many people in a pack at reduced rates. On the contrary, the cost of personal information from the consumer point of view is much higher than from the service provider point of view. A study by Compassed Intelligences surveyed more than 1,000 people in the U.S., U.K. and Canada [20]. The survey asked them to assign a value to their personal information. The results showed that consumers placed an overall value for their information on SNS at between \$62.79 and \$106.40. These studies show that service providers and consumers may have different perspectives regarding personal information.

Other studies that used different methods also had different results. Carrascal et al. [21] studied valuating different types of personal information. They used an auction game based on the reverse second price auction to find the value of personal information and found that the overall median bid value for context-dependent personal information, which is personal information that does not relate to a visited website, was $\overline{X}$ = €7. Some personal information related to an offline identity, such as age, gender and address, had a higher value with an overall median bid of $\overline{X}$ = €25. Moreover, Staiano et al. [22] were inspired by the work of Carrascal et al. and also used a Day Reconstruction Method (DRM) combined with a reverse second price auction mechanism to collect monetary valuation. They reported a different monetary value for personal information from their study, which had an overall median bid value of $\overline{X}$ = €2.

Most studies have different ideas for estimating the value of personal information. Moreover, there are some researches that did not use direct methods, such as surveys or auctions, to ask about the value of personal information. Related intangible incentives were used to estimate the value of personal information. A study by Otsuki and Sonehara used the cost of protecting personal information for each personal attribute by estimating the value of that personal information [23]. Another study developed a tool called "Cloudsweeper," which calculated the value of email accounts using other accounts associated with the email. The emails associated with important account information, such as financial accounts and e-banking accounts, had higher value than the emails that contained only communication information [24]. From the studies and results mentioned above, calculating the exact value of any personal information remains difficult.

Moreover, one of the related works was a study on the ordering of question which aims to improve the motivation for survey participants [25]. The authors organized an experiment which provided two sets of questionnaires, in which each set of the questionnaires asked a set of 33 personal information in difference order. The valuation of each personal attribute was calculated using our techniques, Bayesian probability and graph algorithm. The first set of questionnaires started from personal information with high valuation. The second set of questionnaires, on the contrary, started from low value of personal information to those with higher valuation. The authors found that the participants who received the second set of questionnaire agreed to submit some information at higher percentage (71.42%). The research also studied the time for filling the questionnaires which the second set of participants spend less time in filling the questionnaire. The result of this previous research showed a difference result with our study that the form of requesting environment can affect the result of the study. Using the survey form, participants may review all requested personal attributes in the survey and then select some personal attributes which they feel comfortable to disclose. However, the simulation of the trading platform did not provide any chance for them to review the set of personal attributes. When the participants got the incentive offer, they had to decide without having any hint about the next requested personal attributes and they could not change their decision after they already made a decision.

6. Conclusions

This study aimed to promote the trading activities between service providers' monetary incentives which are offered to consumers and consumers' decisions to disclose their personal information when receiving the service providers' offers. A method was proposed to evaluate personal attributes. Valuations using graph visualization software were made available to service providers.

The evaluations were used to create a questionnaire that motivated consumers to provide more personal information. Results indicated that consumers tended to provide more personal information when the questions were ordered from the most important attribute to the least important. The improvement was more significant when the order was obtained from survey data on participants with the same demographic grouping as the consumers.

Acknowledgments The authors would like to thank Mr. Rathachai Chawuthai, Miss Prariyawan Ratanakrong and anonymous reviewers for several useful advices during this research.

References

[1] Shacklett, M.: Big data wake-up call: Increased online privacy concerns require risk management, available from ⟨http://www.techrepublic.com/article/big-data-wake-up-call-increased-online-privacy-concerns-require-risk-management/⟩ (accessed 2015-03-30).
[2] TRUSTe.: Data privacy is a major concern for consumers, available from ⟨http://www.truste.com/blog/2015/01/28/data-privacy-concern-consumers/⟩ (accessed 2015-03-30).
[3] Hendricks, E.: When your identity is their commodity, available from ⟨http://www.washingtonpost.com/wp-dyn/articles/A9101-2005Mar5.html⟩ (accessed 2015-03-30).
[4] Felt, A. and Evans, D.: Privacy protection for social networking platforms, *Proc. Workshop on Web 2.0 Security and Privacy* (*W2SP '08*) (2008).
[5] Osothongs, A. and Sonehara, N.: A proposal of personal information trading platform (PIT): A fair trading between personal information and incentives, *Proc. Conference on Digital Information and Communication Technology and its Applications* (*DICTAP2014*), pp.269–274, IEEE (2014).
[6] Osothongs, A., Suppakitpaisarn, V. and Sonehara, N.: Evaluating the importance of personal information attributes using graph mining technique, *Proc. International Conference on Ubiquitous Information Management and Communication* (*ACM IMCOM-ICUIMC2015*), 104:1-104:8, ACM (2015).
[7] Osothongs, A., Suppakitpaisarn, V. and Sonehara, N.: A prototype decision support system for privacy-service trading, *Proc. IEEE International Conference on Multimedia Big Data* (*IEEE BigMM2015*), pp.282–283, IEEE (2015).
[8] Osothongs, A., Suppakitpaisarn, V. and Sonehara, N.: A proposed method for personal attributes disclosure valuation: A study on personal attributes disclosure in Thailand, *Proc. 9th International Conference on Information Technology and Electrical Engineering* (*ICITEE 2015*), pp.409–413, IEEE (2015).
[9] Otsuki, M. and Sonehara, N.: Estimating the value of personal information with SNS, utility, *Proc. 8th International Conference on Availability, Reliability and Security* (*ARES2013*), pp.512–516 (2013).
[10] Evans, K.: Personal information in New Zealand: Between a rock and a hard place? *Proc. Interpreting Privacy Principles: Chaos or Consistency? Symposium*, Sydney (2006).
[11] Directive 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, Retrieved November 1, 2014, available from ⟨http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML⟩ (accessed 2015-03-30).
[12] Tene, O. and Polonetsky, J.: Privacy in the age of big data: A time for big decisions, *Stanford Law Review*, No.64, pp.63–69 (2012).
[13] Arvind, N. and Vitaly, S.: Myths and fallacies of personally identifiable information, *Comm. ACM*, Vol.53, No.6, pp.24–26 (2010).
[14] Consumer data collection comes at a cost, available from ⟨http://www.emarketer.com/Article/Consumer-Data-Collection-Comes-Cost/1012634⟩ (accessed 2015-03-30).
[15] Shibchurn, J. and Yan, B.N.: Investigating effects of monetary reward on information disclosure by online social networks users, *Proc. 47th Hawaii International Conference on System Sciences* (*HICSS2014*), IEEE (2014).
[16] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Etienne, L.: Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, Vol.2008, No.10, p.1000 (2010).
[17] Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C.: *Introduction to algorithms*, MIT Press and McGraw-Hill, pp.655–657 (2009).
[18] Welch, B.L.: The generalization of Student's problem when several different population variances are involved, *Biometrika*, Vol.34, No.1, pp.28–35 (1947).
[19] Steel, E., Locke, C., Cadman, E. and Freese, B.: How much is your personal data worth?, available from ⟨http://www.ft.com/intl/cms/s/2/927ca86e-d29b-11e2-88ed-00144feab7de.html⟩ (accessed 2014-11-01).
[20] Burney, K., Brehm, J. and Robinson, K.: Valuing Identity in Today's Digital World: The business case for defining digital identity and how to value it correctly, available from ⟨https://www.unboundid.com/company/news/press/-2013/20130730.php⟩ (accessed 2014-11-01).
[21] Carrascal, J.P., Riederer, C., Erramilli, V., Cherubini, M. and Oliveira, R.: Your browsing behavior for a big mac: Economics of personal information online, *Proc. 22nd International Conference on World Wide Web* (*WWW2013*), pp.189–200 (2013).
[22] Staiano, J., Oliver N., Lepri B., Oliveira, R., Caraviello, M. and Sebe, N.: Money Walks: A Human-Centric Study on the Economics of Personal Mobile Data, *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing 2014, ACM* (*UbiComp2014*), pp.583–594 (2014).
[23] Otsuki, M. and Sonehara, N.: Estimating the Value of Personal Information with SNS, Utility, *Proc. 8th International Conference on Availability, Reliability and Security* (*ARES*), pp.512–516, IEEE (2013).
[24] McCracken, H.: Cloudsweeper's Gmail Security Audit Is Alarming and Useful, available from ⟨http://techland.time.com/2013/06/27/gmail-security/⟩ (accessed 2015-03-30).
[25] Singh, R.K., Suppakitpaisarn, V. and Osothongs, A.: Improving Motivation in Survey Participation by Question Reordering, *Proc. 2016 Pacific Rim Knowledge Acquisition Workshop* (*PKAW 2016*), Lecture Notes in Computer Science, Vol.9806, pp.231–240, Springer (2016).

Ake Osothongs received B.S. in Computer Science from Bangkok University and M.S. in Information Management from Asian Institute of Technology. In 2011, he has started a work in Defence Information and Space Technology Department, Ministry of Defence of Thailand. He was a research assistant in National Institute of Informatics. Now he is a Ph.D. candidate in the Department of Informatics, School of Multidisciplinary in SOKENDAI (Graduate University for Advanced Studies), Japan.

Vorapong Suppakitpaisarn is working as an assistant professor at Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo from May 2015. Before that, he was a project researcher at Global Research Center for Big Data Mathematics, National Institute of Informatics and JST ERATO Kawarabayashi Large Graph Project.

Noboru Sonehara is a professor in the Information and Society Research Division, National Information of Informatics, Tokyo, Japan. He received B.Sc. and M.Sc. degrees from Shinshu University, Japan in 1976 and 1978, respectively. He received Ph.D. from Shinshu University, Japan in 1994. He had been a Director of the Information and Society Research Division since 2006. His current areas of research include ICT security, privacy, trust, risk, resilience, and e-authentication platform design.