Privacy Protection by Anonymizing Based on Status of Provider and Community

Kazutomo Hamamoto, Yasuyuki Tahara, and Akihiko Ohsuga
Graduate School of Information Systems
University of Electro-Communications
Chofu-shi, Tokyo, Japan

***Abstract*— When a user receives personal services from a service provider, the service can be of a higher quality if the user provides more personal information. However, the risk of privacy violation could increase. Therefore, this paper proposes a privacy protection method that realizes avoidance of unwanted information disclosure by controlling disclosable attributes according to the results from monitoring two elements: user background information of the provider and user community status. This is done before disclosing individual attributes corresponding to the privacy policy (i.e., the required anonymity level) by each user. The system architecture based on the aforementioned is also proposed. The validity of the proposed methods was confirmed by a desk model.**

Keywords-Privacy; Anonymity; Lifelog; Personal information

I. INTRODUCTION

When a user receives personal services, such as content recommendation service or action support service, from a service provider by supplying personal information including one's lifelog, preference, age, etc., privacy protection is essential. If the information given to the provider is generally large and detailed, the received service quality increases although the anonymity level lowers. For privacy protection, the offered information should be reduced as much as possible to prevent the anonymity level from lowering. This indicates that a trade-off exists between privacy protection and service quality [1], [3].

When the service quality from the provider is not satisfactory, it is possible to increase it by increasing the offered information quantity, at the cost of lowering the anonymity level. However, in some critical situations, a dilemma whether to increase the information quantity or prioritise anonymity arises. In such cases, the demand of the withdrawal of information that has been disclosed unwillingly is meaningless because it is almost impossible to withdraw or cancel information once disclosed. Thus, it can be said that this trade-off control is one-way and has so-called 'No Entry Area', as shown in Figure 1, namely in that area it is impossible to increase the anonymity level once lowered such as from (1) to (2).

Considering these circumstances, this paper proposes a method that realizes avoidance of unwanted information disclosure by controlling openable attributes (i.e., the attributes disclosable as per required anonymity level) according to the results from monitoring two elements: the user background information of the provider (i.e., the information that the service provider already possesses about the user) and the user community status (i.e., head count etc. of the community including the user) that influences the anonymity level. This monitoring is done before disclosing individual attributes corresponding to the privacy policy (hereafter the required anonymity level) set by each user. This paper aims to propose such privacy protection methods to enable service acquisition corresponding to the offered information without any unintended personal information leakage.

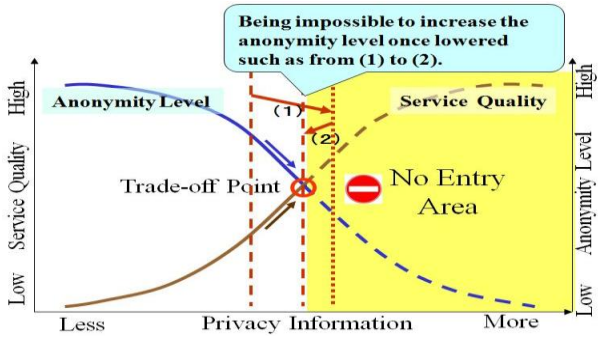


Figure 1. One-way Trade-off Control

This paper is organized as follows: in Section Ⅱ, the related works are described, and different problems are analysed in Section Ⅲ. The proposed methods are described in Section Ⅳ. In Section Ⅴ, evaluation and validity verification are discussed. In Section Ⅵ, the architecture is described, followed by an outline of future works in Section Ⅶ. Finally, Section Ⅷ concludes this paper.

II. RELATED WORKS

Relevant works [1]-[15] were analysed on the basis of the following three viewpoints.

A. Storage Location of Personal Information and Disclosing Condition

In many cases, personal information of each user is stored in a secure server located between users and providers, and then batch processing, such as data analysis or anonymization, is carried out in bulk [4], [5], [9], [12], [13]; this architecture concept is adopted in this paper. The privacy policy that in-

DOI: 10.5176_2010-2283_2.1.146

cludes the purpose of the use of personal information or the required anonymity level (i.e., identifying probability) is collated from users and providers, and on satisfying certain conditions, personal information is disclosed [3], [4], [5], [10], [11], [14]; disclosing condition of this paper is original such as to be proposed afterwards.

B. *Privacy Protection methods and Specific individual identification*

For policies having different aforementioned collations, it is impossible to disclose information, although some ways such as obeying the dictates of the user are taken to make progress [1], [2], [6], [7], [12], [13]. Techniques that make information granularity blunt or rough, except the K-anonymity method [10] are used to manage anonymity [3], [5]; the concept of K-anonymity method is adopted in this paper. Specific individual identification is possible in systems that handle information such as name [1], [8], [12], [14]; however, in many cases it is impossible because of various privacy protection techniques [2], [7]. It is possible to identify a unique individual by using a technique called 'shadow attack' that monitors users of the services provided from the server [11]; to prevent specific individual identification by privacy protection is the final goal of this paper.

C. *Explicit trade-off control*

Although almost all related works do not refer to the trade-off between anonymity and service, some works explicitly consider it. One trade-off is the balance between the received services and the offered attributes achieved via user hands-on control [1], [4], [6], and the other is a system that searches for the type and granularity of openable attributes automatically by using the machine learning technique [3]. The user's load cannot be neglected in the former, and in the latter some services cannot be utilized. A research on the trade-off between privacy and trust [15] suggests the presence of inherent affinity between trust and service.

Although the aforementioned works discuss the trade-off, very few describe user background information and user community status that influences the anonymity level. However, when considering that various new services and applications that utilize lifelogs collected from blogs or social network services spread and circulate in the network, privacy protection is important. This is achieved by careful control of the anonymity level and the disclosable attributes from the trade-off standpoint, and that is the aim of this paper.

III. PROBLEM ANALYSIS

A. *Prerequisite Framework*

The prerequisite framework describing the two elements: the user background information of the provider and the user community status, can be considered to be similar to that found in the literature; this framework includes a secure server between users and providers that stores personal information and performs various processes such as anonymization. The secure server is called the privacy protection server (PPS) in this paper and performs maximum privacy protection on the basis of careful trade-off control by considering the two elements. It uses the concept of K-anonymity [10] that ensures anonymity by controlling the number of people having the same attribute that is more than the K constant. Figure 2 shows this framework with a series of processes from (1) Personal Information and Required Anonymity Level to (4) Service Contents. Relevant specific architecture is described in Section VI.

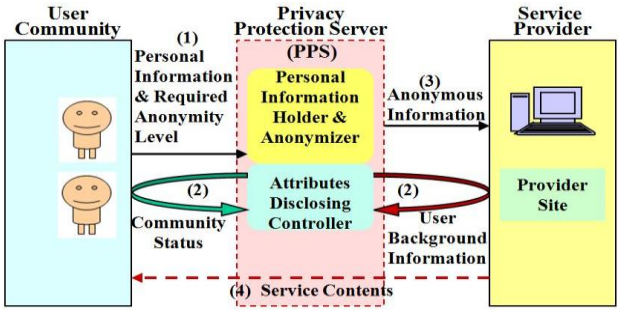


Figure 2. Prerequisite Condition Framework

B. *Two problems*

1) *Problem1: Dealing with user background information*

Here, let numerical value L be the required anonymity level, a set of attributes R be the openable attributes corresponding to L and function f (L) = R. In order to avoid careless entry of unintended information disclosure in the 'No Entry Area' shown in Figure 1, it is possible, without any careless disclosure of all R, to gradually increase the openable attributes from R´ (f (L + α) = R´) to R (#R = #R´ + β) by lowering the anonymity level stepwise from L + α to L with monitoring the service quality [18]. Here, α and β are positive integers.

If the provider has a set of attributes M as the user background information, by combining M with R´, the substantial attributes disclosed to the provider will be M + R´. This indicates that it is possible for the substantial anonymity level to lower below L, and consequently such unintended disclosure could happen. Figure 3 shows these aspects. To this end, it is necessary to search for the user background information of the provider before disclosing the information and take some appropriate measures according to the results.

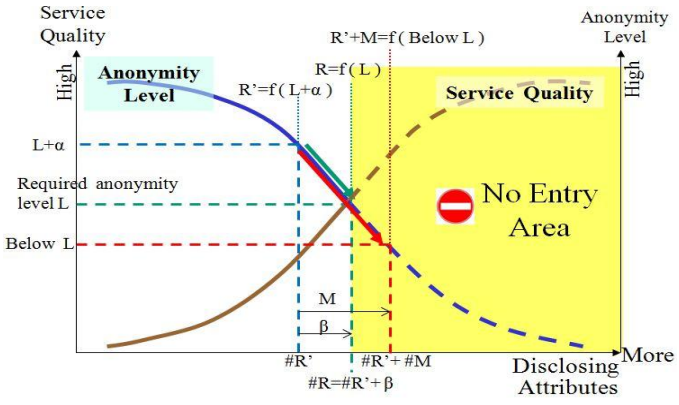


Figure 3. How to deal with the openable attributes

2) Problem2: Dealing with community status

It is unavoidable that cases such as the required anonymity level cannot be defended, occur because of changes in the community status as long as the anonymity level is based on K-anonymity. Even if a moving average or some regularized indicator is applied to reduce the influence of the change, it is not completely eliminated.

Therefore, in this paper, disclosing control of the openable attributes is performed along with informing the user about the openable attributes in advance by sensing such changes. Thus, it is possible to reflect the user's intention in disclosing control of attributes beforehand. It resembles an advanced demand signals scheme (ADS) [19] that controls the signals beforehand by measuring traffic flow towards the intersection. Figure 4 shows these situation.

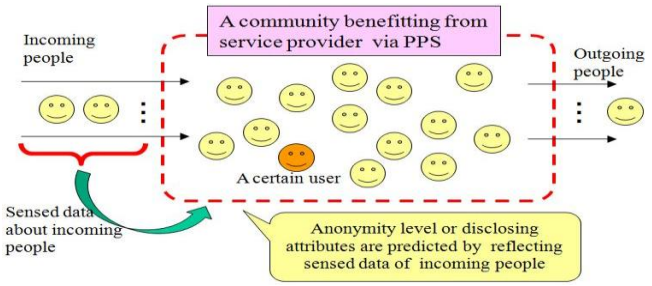


Figure 4. How to deal with community change

IV. PROPOSED METHODS

A. Anonymizing and disclosing control

The K value cannot be used as the anonymity level for each user because it changes according to the community scale. Therefore, the required anonymity level L is introduced that has four levels that decrease in the order of 3, 2, 1 and 0. The value of K corresponding to each level of L is appropriately determined depending on the community scale. The PPS shown in Figure 2 determines the openable attributes according to L of each user by monitoring consolidated personal information, and then discloses them to providers after proper processes as proposed afterwards. The decision tree learning algorithm is applied to control the disclosing attribute group and order [4].

In general, a selective attribute in the decision tree learning algorithm is identified for efficient classification and fast access to the target object by choosing attributes such as information gain that generates big entropy. However, on the contrary, in this paper, an inefficient decision tree is generated for the unintended disclosure of the target object (i.e., privacy protection object) by using such attribute because the entropy is comparatively low; as a result, by using this tree, privacy is protected. Hereafter, a specific case is illustrated.

Table Ⅰ shows the personal information of a group of people in an airport waiting room. The table shows the entropy when classified according to each attribute value. For the case of Alice, when classified by gender, the entropy is calculated from the definition referring to Figure 5 (a) as follows:

$$-\sum_{i=1}^{k} p_i \log p_i = -\{2(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2})+4(\frac{4}{4}\log\frac{4}{4})\}/6 \cong 0.33$$

TABLE I. PRIVACY INFORMATION OF A GROUP

Name	Att 1	Att 2	Att 3	Att 4	Att 5
	Gender	Job	Blood	First Trip	Goal
Alice	F,0.33	Stu,0.53	A,0.46	Yes,0.46	USA,0.33
Bob	M,0.53	Stu,0.53	AB,0	Yes,0.46	USA,0.33
Mike	M,0.53	Busi,0.33	A,0.46	No,0.46	UK,0
John	M,0.53	Busi,0.33	O,0	No,0.46	Fra,0.33
Hanak	F,0.33	Stu,0.53	A,0.46	No,0.46	Ita,0
MikeJ	M,0.53	Stu,0.53	B,0	Yes,0.46	Fra,0.33

(Notes) Att: Attribute, F: Female, M: Male, Stu: Student, Busi: Business, USA: United States of America, UK: United Kingdom, Fra: France, Ita: Italy

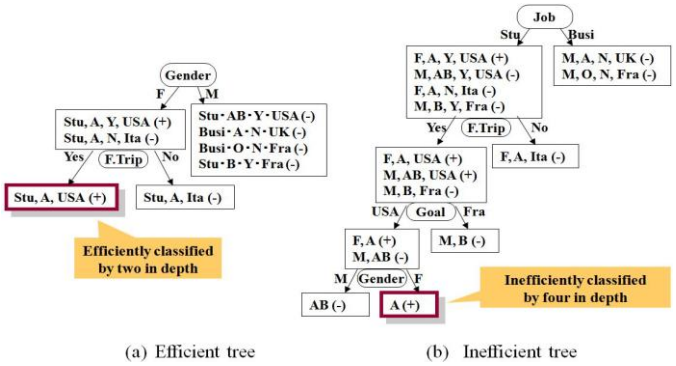


Figure 5. Decision Trees of Alice

Thus, as shown in Figure 5, although the efficient decision tree needs only two in-depth attributes to specify the object, the inefficient tree needs four in-depth attributes for the same. The inefficient classifying tree is used for privacy protection. The attributes are individually disclosed from such an attribute because the object is not easily specified. In particular 'Job: Student → First Trip: Yes → Goal: USA' becomes a disclosing order. However, in order to avoid complex processes all openable attributes determined by L are disclosed simultaneously not individually in the following section.

B. Proposed Method 1 (For Status of Service Provider)

1) Technique to avoid the influence of user background information

Figure 6 shows a situation where the provider has some user background information. The user discloses the attributes 'a' and 'b' that are determined to be openable by the required anonymity level L = 2. If the provider already has the user background information equivalent to attribute 'c', the substantial anonymity level lowers from L = 2 to L = 1 by combining 'a' and 'b' to 'c'.

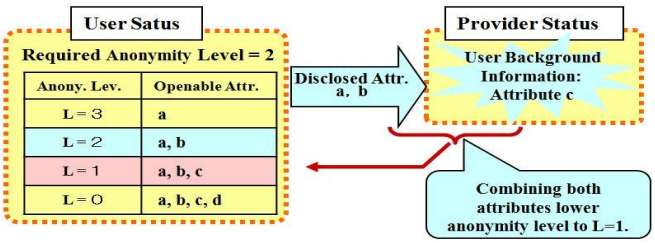


Figure 6. Influence of user background information

Using Table Ⅰ, the following situation can be assumed.

For Alice, the openable attributes corresponding to the required anonymity level L = 2, indicating K = 3, are 'Job: Student,' 'Blood: A' or 'First Trip: Yes', and those are disclosed to the provider concerned. It is located in the airport, and so possible to acquire beforehand the information such as boarding members list including name, nationality, destination and first-time overseas travellers etc.; accordingly the provider can select three people Alice, Bob and MikeJ when received the 'First Trip: Yes' attribute. Under the circumstances, if the provider acquires information that MikeJ was in a French embassy three hours ago from a location service provider, MikeJ's 'Goal' is possibly France. Thus, the target person having attributes such as 'Job: Student,' 'Blood: A' or 'First Trip: Yes' is Alice or Bob, which indicates that the substantial anonymity level lowers from the initial level L = 2 (K = 3) to L = 1 (K = 2), then the required anonymity level cannot be defended .

In order to avoid this, this paper proposes a certain method (*Proposed method 1*) that appropriately changes disclosing control of attributes according to the results obtained from monitoring the user background information before disclosing openable attributes corresponding to the required anonymity level. Contents and procedures are shown in each step in Figure 7, illustrating the conditions of the selected sample as follows together with Table Ⅰ or Figure 6:

In step (1), the candidate provider is appropriately searched for by the query words composed of attributes common to all users. An example of a query is '(Gender: Male, Female), (Job: Student, Business), (Blood: A, AB, O, B), (First Trip: Yes, No), (Goal: USA, UK, France, Italy)' like the LCM (least common multiple).

In step (2), each user selects several providers, each denoted by H, of his/her choice among the providers that were searched for by step (1).

In step (3), the appropriate pages of each selected provider site are downloaded and are to be searched for the same attributes of each user from the viewpoint of how much of each user's attribute is included in the pages. For the case of Mike, instance of such attributes is 'Gender: Male, Job: Business, Blood: A, First Trip: No, Goal: UK'.

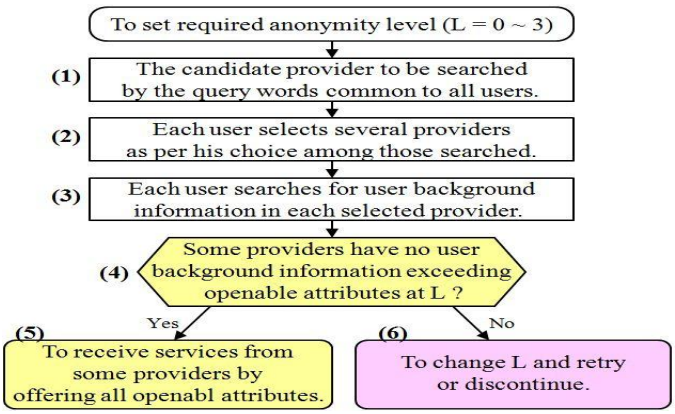


Figure 7. User Background Information Search Flow

In step (4) and (5), the service is received by disclosing openable attributes corresponding to the required anonymity level L from the provider that is considered not to have attributes exceeding the range of the openable attributes. For example, if L = 2 in Figure 4, only the provider that does not have 'c' and 'd' attributes will be able to provide some services.

In step (6), as for the other providers, it is possible that they may not maintain the L level when disclosing openable attributes, therefore each user retries after changing the L level, or the service itself should be cancelled; this is one process for the trade-off between privacy protection and service quality.

2) Technique to search for user backgroud information

Considering all attributes of each user that have already been registered in PPS, the target is to find the number of the same attributes in the provider site. In particular, by using the pair of words of the attributes and its values, the co-occurrence frequency of the pair is measured. As for the co-occurrence level of a pair of words, two typical coefficients are used as the index of relativity between the pair: the Jaccard coefficient and the Simpson coefficient [16]. In this paper, the Jaccard coefficient is used considering that the paired words have a tendency to appear simultaneously and it is important to determine whether the target site also has this tendency. Jaccard coefficient functions properly in such case.

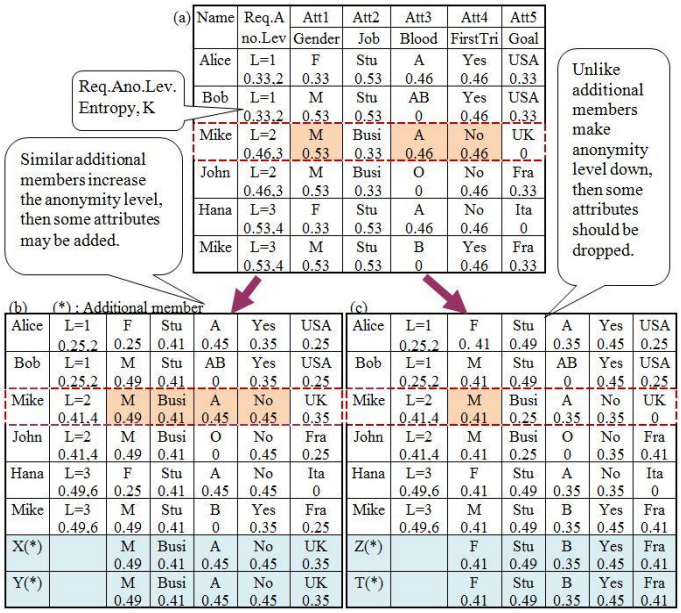
If the number of such pairs offered from the user and the chosen providers in step (2) in Figure 5 are assumed to be P and H, respectively, co-occurrence level $\boldsymbol{J}$ (the Jaccard coefficient) of each pair of words is shown as a two-dimensional array. Assuming P and H for the first and second dimensions, respectively, $\boldsymbol{J}$ is defined as follows.

$$\boldsymbol{J} = [[J_{11}, J_{12}, \ldots, J_{1p}], [J_{21}, J_{22}, \ldots, J_{2p}], \ldots, [J_{h1}, J_{h2}, \ldots, J_{hp}]]$$

Among P pairs, if the number of the pair that can be disclosed by level L is assumed to be R, the number of the pair that cannot be disclosed is Q = P − R. For a certain provider, if any J_{hp} value of each of the Q pairs does not exceed a certain threshold level T, such a provider can be considered not to have user background information about Q, and then the service can be received by only offering the attributes R, openable corresponding to level L, as described in step (4) and (5). Conversely, if any J_{hp} value of each of the Q pairs exceeds the threshold level, the service cannot be received from such provider.

C. Proposed Method 2 (For Status of User Community)

As stated in Section Ⅲ, the required anonymity level L cannot be defended because of the community status change. Therefore, to defend L, it is necessary to control the disclosing attributes according to the current anonymity level. Figure 8 shows this situation considering the case of Mike as an example. Figure 6 (a) shows the initial situation in which Mike has L = 2 (K = 3) and the three attributes, 'Male', 'A' and 'No' can be disclosed. Figure 6 (b) shows that the anonymity level increases because two members having same attributes as those of Mike join the group; thus, another attribute 'Busi' can be disclosed. Figure 6 (c) shows the opposite situation; therefore,



(a) Name	Req.A no.Lev	Att1 Gender	Att2 Job	Att3 Blood	Att4 FirstTri	Att5 Goal
Alice	L=1 0.33,2	F 0.33	Stu 0.53	A 0.46	Yes 0.46	USA 0.33
Bob	L=1 0.33,2	M 0.53	Stu 0.53	AB 0	Yes 0.46	USA 0.33
Mike	L=2 0.46,3	M 0.53	Busi 0.33	A 0.46	No 0.46	UK 0
John	L=2 0.46,3	M 0.53	Busi 0.33	O 0	No 0.46	Fra 0.33
Hana	L=3 0.53,4	F 0.33	Stu 0.53	A 0.46	No 0.46	Ita 0
Mike	L=3 0.53,4	M 0.53	Stu 0.53	B 0	Yes 0.46	Fra 0.33

(b)						
Alice	L=1 0.25,2	F 0.25	Stu 0.41	A 0.45	Yes 0.35	USA 0.25
Bob	L=1 0.25,2	M 0.49	Stu 0.41	AB 0	Yes 0.35	USA 0.25
Mike	L=2 0.41,4	M 0.49	Busi 0.41	A 0.45	No 0.45	UK 0.35
John	L=2 0.41,4	M 0.49	Busi 0.41	O 0	No 0.45	Fra 0.25
Hana	L=3 0.49,6	F 0.25	Stu 0.41	A 0.45	No 0.45	Ita 0
Mike	L=3 0.49,6	M 0.49	Stu 0.41	B 0	Yes 0.35	Fra 0.25
X(*)		M 0.49	Busi 0.41	A 0.45	No 0.45	UK 0.35
Y(*)		M 0.49	Busi 0.41	A 0.45	No 0.45	UK 0.35

(c)						
Alice	L=1 0.25,2	F 0. 41	Stu 0.49	A 0.35	Yes 0.45	USA 0.25
Bob	L=1 0.25,2	M 0.41	Stu 0.49	AB 0	Yes 0.45	USA 0.25
Mike	L=2 0.41,4	M 0.41	Busi 0.25	A 0.35	No 0.35	UK 0
John	L=2 0.41,4	M 0.41	Busi 0.25	O 0	No 0.35	Fra 0.41
Hana	L=3 0.49,6	F 0.41	Stu 0.49	A 0.35	No 0.35	Ita 0
Mike	L=3 0.49,6	M 0.41	Stu 0.49	B 0.35	Yes 0.45	Fra 0.41
Z(*)		F 0.41	Stu 0.49	B 0.35	Yes 0.45	Fra 0.41
T(*)		F 0.41	Stu 0.49	B 0.35	Yes 0.45	Fra 0.41

Figure 8. Changes of Anonymity

the openable attribute should only be limited to 'Male' to defend L.

In order to simplify the influence of community change, there are some methods such as using regularized entropy in the anonymity level calculation [4] or a certain attribute disclosing index considering lifelog statistics and community character. In spite of these methods, the influence cannot be easily simplified, and thus cannot be considered as a suitable solution.

We propose an advance agreement attribute disclosing controlling method (*Proposed method 2*), namely reflecting the user's intention in advance. It is outlined as follows: the method senses momentarily information in advance such as the number of people moving toward the service area or the attributes associated with those people; forwards those sensed data to the server (PPS) in the service area whenever just after sensed; forecasts instantly the anonymity level change and the openable attributes by analyzing those gathered data; listens to the user's intention beforehand about attributes to be disclosed; and determines whether to disclose additional attributes by obtaining user consent. Figure 9 illustrates such essence as the described above.Thus, appropriate real time disclosing control can be performed according to the current conditions. Details are shown in Section Ⅴ by using a simulator.

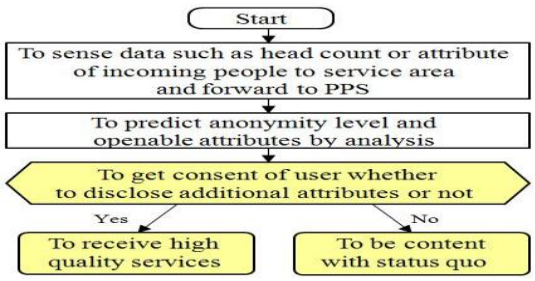


Figure 9. Predicting process Flow

V. Evaluation and Validity Verification

A. *Evaluation and Validity forProposed Method 1*

By using the case group in Table Ⅰ, it is verified that the proposed method 1, described in Section Ⅳ- B (Figure 7), is appropriate when treating user background information and anonymity level, and disclosing attributes from two evaluation aspects: (#1) There is no risk in disclosing unintended personal information, (#2) There is no sacrifice of the unintended anonymity level.

In step (1) in Figure 7, the query words used are common to all users, so the anonymity level is the highest and each user consents the disclosure of such personal information beforehand by understanding the purpose of the use. Thus, aspects (#1) and (#2) can be realized. In steps (2) and (3), while processing only one-sided download of the contents of the related page of the provider site takes place; thus, aspects (#1) and (#2) can be realized. In steps (4) and (5), only when any J_{hp} value of each of the Q pairs does not exceed threshold T, all openable attributes R corresponding to L are disclosed and receive the service; thus, aspects (#1) and (#2) can be realized. In step (6), the processing is performed from the user's standpoint; therefore, aspects (#1) and (#2) can be realized. Thus, the proposed method 1 is verified to be appropriate from both evaluation aspects.

Table Ⅱ shows an example of the result of the provider searching experiment for user background information in step (3) assuming the case of Mike in which L = 2 (K = 3). Here, the sites hit by Google are considered as providers. If the threshold T is set to be 0.1, because any attribute of the Q pairs (non-marked columns) of any provider does not exceed T , the services can be received from all three providers. Practically, such threshold level should be carefully determined based on various system conditions.

TABLE II. JACCARD COEFFICIENT VALUE OF EACH ATTRIBUTE

Provider	Attr1	Attr 2	Attr 3	Attr 4	Attr 5
	Gender	Job	Blood	FirstTrip	Goal
	Male	Business	A type	No	UK
(a)	0/4 (0)	2/22 (0.09)	2/7 (0.29)	0/1 (0)	0/0 (0)
(b)	0/0 (0)	0/0 (0)	1/1 (1)	0/1 (0)	0/7 (0)
(c)	0/21 (0)	0/6 (0)	2/12 (0.17)	0/16 (0)	0/2 (0)

- Figures show Jaccard Coefficient ($|A \cap B|$) / $\sum$ ($|A \cup B|$).
- Marked cells show openable attributes corresponding to Required Anonymity Level L = 2 (K = 3).

B. *Evaluation and Validity forProposed Method 2*

As in the previous section, it is verified that the proposed method 2, described in Section Ⅳ- C, is appropriate by using a multi-agent simulator [17], [20]. In particular, the possibility of real time disclosing control based on an advance agreement attribute is considered. That is, acquiring personal information of the people moving towards the service area beforehand, informing the user in the service area of the change of openable attributes and negotiating with the user as to which attribute is to be opened with high priority may be achieved. Figure 10 and Figure 11 show the simulation status.

In Figure 10, pedestrian agents assuming customer move toward the boarding gate in the airport where after passing A area receive services at B area (service area) and pass C area toward exit. Various sensing are performed at A area. Key parameters are A_t and B_t, that are the average time for an agent to pass the A area and the average time to receive the service at the B area, respectively, unit time being a step. Other parameters are commonsensibly and properly decided by the simulator. The value of the anonymity level is defined to be the number of agents in the A or B area having the same attribute as the marked agent (user) locating in the B area to receive services.

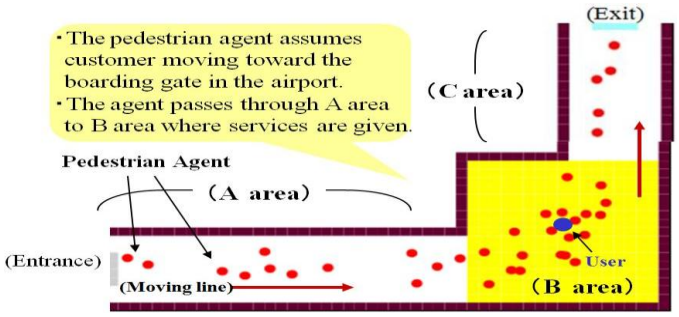


Figure 10. Simulation of Community Status Change

Figure 11 shows the simulation by assuming A_t = 40 steps and B_t = 80 steps; therefore, α = 2 such that $B_t = \alpha \times A_t$. The upper and lower graphs show the situations of the A and Bareas, respectively. A(t), A1(t) and A2(t) show the number of people, the anonymity level of attribute 1 and the anonymity level of attribute 2, respectively, at Time = t steps in the A area. Similarly, B(t), B1(t) and B2(t) correspond to the number of people, the anonymity level of attribute 1 and the anonymity level of attribute 2, respectively, in the B area. From the graph, B(120) ≒ A(40) + A(80), B(160) ≒ A(80) + A(120), ... or B1(440) ≒ A1(360) + A1(400), B1(480) ≒ A1(400) + A1(440), ... can be read, which suggests that the situation in B area can be predicted from the situation in the A area. This is because B (40n) ≒ A (40(n-1)) + A (40(n-2)) can be obtained by the average value because of $B_t = 2 \times A_t$, where n is a positive integer.

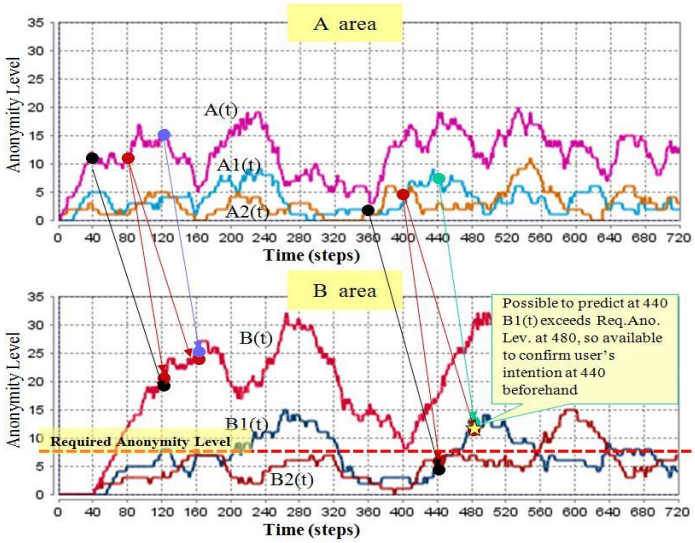


Figure 11. Simulation of Anonymity Change

That is, the anonymity level of B1(t) in Figure 11 shows that it can be predicted at Time = 440 steps that the value of B1 at Time = 480 steps exceeds the required anonymity level beforehand. Then, at Time = 440 steps, it is possible for the user to negotiate and decide whether to disclose additional attributes; thus, an advance agreement disclosing control is possible. This suggests that it is verified from the evaluation aspect (#1) viewpoint. In contrast, in the situation in which the number of already disclosed attributes should be decreased by forecasting, it is possible to discontinue disclosing as soon as possible by prior notification, and thus the sacrifice of the anonymity level can be minimized as per the user's consent. Therefore, aspect (#2) can be achieved.

C. *Evaluation of the Proposed Methods for Trade-Off*

The proposed methods should be evaluated from the viewpoint of the contribution to an effective trade-off between service and anonymity from the following two aspects: (a) easy and rapid acquiring trade-off point (i.e., the balancing point between service and anonymity) and (b) no unintended information disclosure and availability of a slightly higher service level.

First, this paper aims to introduce the required anonymity level (L) that can be set by each user and notify such users of the openable attributes determined by L; thus, the mechanism that a phased adjustment is enabled is considered. Moreover, although it is usual to individually disclose openable attributes in order [3], [18], in this paper, all openable attributes are disclosed simultaneously; therefore, the balance point can be efficiently attained. This suggests that the evaluation of aspect (a) is clarified. Second, unintended information disclosure can be avoided by the proposed method 1 in case the provider has some user background information; it is possible to raise the service level by increasing information disclosure intentionally by the proposed method 2 in case the community status changes. This indicates that the evaluation aspect (b) is also clarified.

VI. ARCHITECTUR

The components that compose PPS are shown in Figure 12.

(1) In the 'Personal Information Holder' module, personal information etc. offered by each user are kept in the holder as a database and managed collectively.

(2) In the 'Query Generator and Provider's URL Retriever' module, query words common to all users are composed and supplied to the network to retrieve candidate providers.

(3) The 'User Background Information Retriever' module retrieves the contents of the URL's appropriate pages.

(4) In the 'Openable Attributes Analysis' module, both the openable and non openable attributes are determined provisionally according to the required anonymity level for each user by analyzing the database.

(5) In the 'Openable Attributes Adjusting' module, the anonymity level is calculated every Δt repeatedly reflecting present community status and the openable attributes determined at (4) are adjusted.

(6) In the 'Openable Attributes Fixing' module, the openable (or non openable) attributes that were adjusted in (5) are approved by users through negotiation, accordingly fixed.

(7) In the 'Jaccard-Coefficient Measure and Check' module, the similarity between non openable attributes fixed in (6) and the equivalent contents of the pages acquired in (3) is measured by Jaccard coefficient, and if judged user background information exist, then some alarm is given to the user.

(8) In the 'Service Retriever' module, all the openable attributes are disclosed when there is no user background information in (7) step and proper service contents are retrieved.

(9) The 'Obscuring and Abstracting Process' module, although not illustrated in PPS in Figure 2, performs obscuring and abstracting properly when such functions are desired to avoid privacy violation, although it is future task.

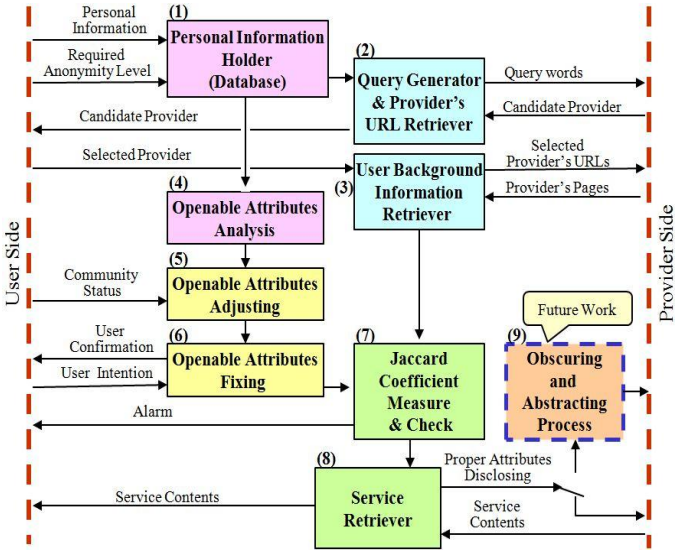


Figure 12. Components of PPS architecture

VII. Future Works

Some experimental evaluating tests for validity verification and specific architecture need further investigation. Especially, when those two proposed methods are actually applied, it is future work to pursue how to combine and coordinate them considering various system conditions.

In the proposed method 1, the threshold used for judging user background information and the technique choosing the most appropriate page of the provider site are the key problems to be pursued and should be solved in the future. In the proposed method 2 simulation, although the airport was assumed for example, given some conditions defining α and A_t, the application of this method becomes possible in locations such as a shopping street or an amusement park. However, there are problems because of the differences between simulation and reality or the limitation of the simulation itself, and this is also an area for future work.

Eventually, when lifelogs increase in number, privacy violation by the secondary use of other providers becomes a big problem. Figure 13 shows an example of privacy violation by such secondary use; that is, if some lifelogs acquired by a separate provider are linked and analysed, the individual image becomes clearer. A key to prevent such analysis is, when personal information is offered from one provider to another, to ensure the anonymity level required by the original owner of the information based on the idea of the proposed methods; however, it is a future task.

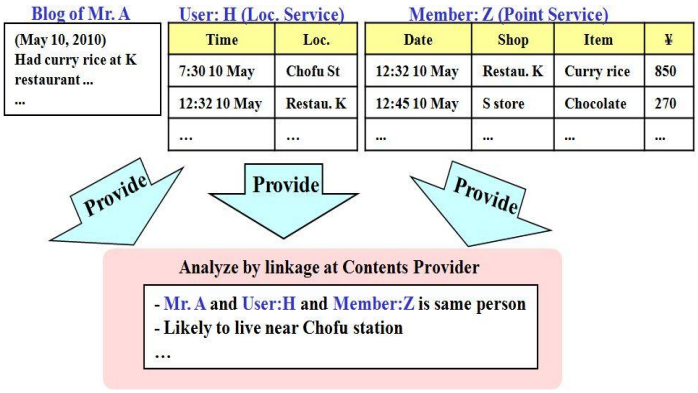


Figure 13. Inference by Linkage Analysis

VIII. Conclutions

The trade-off existing between privacy protection and service quality was discussed. Two problems associated with appropriate trade-off control, namely those with user background information of the provider and community status, were treated and two counter-measures were proposed. Careful anonymity level control through phased adjustment considering user background information and advance agreement attribute disclosing control considering community status are crucial to prevent unintended privacy information leakage.

References

[1] T.Yamabe, K.Fujinami, T.Syoji, N.Nakamura, T, Nakajima, "PENATES:Privacy protection architecture for context-aware environments," in Computer Symposium, pp.55-64, Tokyo, Nov 2004

[2] S.Tamaru, A.Iwatani, K.Takashio, H.Tokuda, "Application framework for personalized public space considering privacy," IPSJ report, Japan, pp.49-56, 2003-OS-93

[3] T.miyamoto, T.Takeuti, T.Okuda, K.Haruki, Y.Ariyoshik, S.Simojo, "Proposal for profile control mechanism considering privacy and quality of personalization services," DEWS2005, Japan, 6-A-01

[4] M.Imada, K.Takasugi, M.Ohta, K.Koyanagi, "LOOM:Loosly managed privacy protection method for ubiquitous networking environments," IEICE journal B, Japan, Vol.J88-B No.3, pp.563-573, 2005

[5] K.Nakanishi, K.Takashio, H.Tokuda, "A concept of location anonymization," IPSJ journal, Japn, vol.46 No.9, pp.2260-2268, 2005

[6] N.Hirotaka, N.Nobuhiro, "Service platform for privacy Controllable Tag," IPSJ report, Japan, 2007-UBI-16, pp.57-63, 2007

[7] T.Sanda, S.Yamada, E.Kamioka, "Proposal for a method of privacy protection in ubiquitous computing environments," IPSJ journal, Japan, 2003-MLB-26, pp.45-51, 2003

[8] M.Langheinrich, "A privacy awareness system for ubiquitous computing environments," Proc.UbiComp, Vol.2498, Springer-Verlag, pp.237-245, 2002

[9] Ginger Myles, Adrian Friday, N.Davies, "Preserving Privacy in environments with location-based Applications," IEEE Pervasive Computing, Vol.2, No.1, pp.56-64, 2003

[10] L.Sweeney, "K-anonymity:a model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-based System, 10(5), pp.557-570, 2002

[11] Linda Pareschi, D.Riboni, C.Bettini, "Protecting users' anonymity in pervasive computing environments," Proceedings of the Sixth Annual IEEE International Conf. on PerCom and Communications, pp.11-19, 2008

[12] Jason I.Hong, James A. Landy, "An architecture for privacy-sensitive ubiquitous computing," Proceedings of the 2nd international conference on Mobile systems, applications and services, pp.177-189, 2004

[13] K.Sato, "Life Log : About the Profit Use of the Cellular Phone Behavional Data that Considers the Privacy Protection," IPSJ magazine, Japan, Vol.50 No.7, pp.598-602, July 2009

[14] P3P, http://www.w3.org/P3P/

[15] Jean-M.Seigneur, C.D.Jensen, "Trading Privacy for Trust," Trust Management LNCS, Vol. 2995/2004, 93–107, 2004

[16] Y.Matsuo, H.Tomobe, K.Hashida, H.Nakajima, M.Ishizuka, "Social Network Extraction from the Web information," JSAI journal, Japan, Vol.20 1E, pp.46-56, 2005

[17] ArtiSoc, http://mas.kke.co.jp/

[18] K.Hamamoto, Y.Tahara, A.Ohsuga, "Proposal for Profile Opening Method Considering Privacy by Anonymization," JWEIN10 symposium, Japan, Proceeding 10, August 2010

[19] Y.Kato, T.Hasegawa, "Effect of Advanced Demand Signals scheme" Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th 708-712, 2002

[20] T.Kaneda, KK.Enginering, Nagoya Ins.of Tech. Univ., "Pedestrian Simulation by Artisoc," Book published in Japan, ISBN978-4-904701-17-1, pp.79-114, 2010