ANALYZING THE ATTRIBUTES OF PERSONALIZATION INFORMATION AFFECTING STORAGE LOCATION

Pekka Jäppinen
Lappeenranta University of Technology
P.O.Box 21, 53850 Lappeenranta, Finland

Jari Porras
Lappeenranta University of Technology
P.O.Box 21, 53850 Lappeenranta, Finland

ABSTRACT

In this paper the properties of different type of personal information used for providing personalized services are studied. Different dependencies of the information as well as the ways of using that information are considered. The objectives of the study are to minimize the amount of redundant personal information and to increase the usability of services. As a result the attributes of personalization information are reflected to the proper location of the information.

KEYWORDS

Personalization, personal information, information location, information dependencies, privacy.

1. INTRODUCTION

The past few years have shown us the rapid development both in the Internet and in the mobile communication services. At the same time the amount of users of electronic services has been growing. This means that the preferences on service usage and type as well as accessing equipment and method vary a lot. Therefore, techniques allowing identification of the valuable services and content as well as customizability are required. Personalization is a technique that allows this.

Personalization can be seen as a technique that allows targeting the requested information or services according to the user's preferences. Personalization is also an effort where the company offering services and the customer using those services has to cooperate. Customer has to provide the necessary information and the company has to use it for providing targeted services. By giving up the asked information the customer relies on the company and its ability and intentions to use the information only for the favor of the customer. For the user point of view this means increased risk, as the customer loses the control over his own data. For company point of view the customer data is needed for making the personalization possible but also for increasing the customer satisfaction and thus the loyalty of the customer. Therefore personalization is beneficial for both parties and will be one of the key aspects of future services. In order to increase the usability of services the location of the necessary information should carefully be considered.

Personalization is not only very important aspect but also very challenging as several factors affect the final result. The balance between usability and privacy has to be taken into account [MAB][VOT]. In this paper the underlying structures of personalization are studied. Based on studies of existing services using personalization both the properties of the required personalization information as well as the use of this information is considered. All the factors are related to the storage location of the personalization information. By doing so the amount of redundant personal information is minimized.

2. PLACES TO STORE THE INFORMATION

Personalization typically consists of two players, user and the service provider. Therefore, the logical place to store the information is either at the *user* or at the *service provider* end. In currently used personalization approaches the information is mostly stored at the service databases from where it is used whenever needed. This however is not always the most optimal place for the information as shown later on. The other logical possibility would be to store the information at the user's end at some device from where it could be easily modified. A third viable possibility is to have *trusted third party* that holds the personalization information and provides it for the service providers. All the cases are shortly analyzed below

Service provider: This is the traditional place to keep the customer information, as it does not require any special equipment at the user end. Service provider holds huge databases containing different information about the users, such as name, address, phone numbers, etc. After the information is gathered it is accessible by the service all the time, but only for the services using the given database. This usually means that when joining to a new service the user has to give the personal information in order to get personalized services. Figure 1 presents a case where user registers into two different services and thus provides the information for both service providers.

1st time service request + personal information   Service

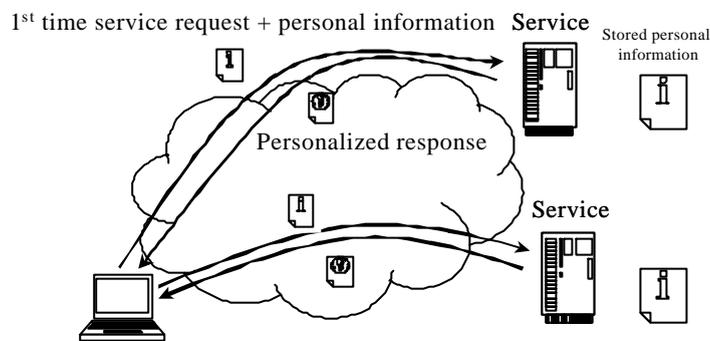Stored personal information

Personalized response

Service

Figure 1. Service provider based approach

Drawback of this method is that several services require the same information. Thus user has to provide the information for all the services separately. This applies especially to mobile users who might need to use various location or service provider dependent mobile services. Another big issue is the accuracy of the information. When the same information is stored on many different places, it is hard for the user to keep it updated at everyplace. Therefore, it is very likely that most of the services eventually hold outdated information. It is also good to remember that most of the credit card numbers have been stolen from databases of web stores instead of the transmission path. [SCH]

Third party: In order to reduce the redundant information a third party approach could be used. In this approach a well known third party can hold the personalization information, from where it can be accessed by various services. For user this means that he has to update his information only at one place. This will make it more likely that the information stays accurate. Figure 2 presents this approach. As user requests service the service provider and third party negotiate the necessary personalization parameters. As a result user receives personalized response.

While the personalization information is at one place, the third party is the single weak link of the system. If the connection to the third party server breaks, there is no access to the personalization information. There is no way for the user or the service provider to affect on the properties of the third party server. On security point of view there is only a single place where the information is stored and it is very likely that a third party that provides such a personalization service takes extra care of the information it stores. This might not be in the case of all service providers.

There exist two single sign-in (SSI) architectures, i.e. Microsoft .NET passport [PAS] and Liberty Alliance [LIB]. Currently .NET passport requires the service provider to pay a fee to Microsoft. This means that small services are not going to adopt the passport. It is also unlikely that conference arrangers are willing to do the extra work required to join in the SSI system especially when all they need is simple registration information for one time event. In Liberty architecture service and identity providers form circles of trust, in which participants transfer information about the user. The creation of a circle of trust requires negotiations between the participating partners. Like in .NET passport this is not suitable for one time services provided by small service providers. Therefore, a simpler solution is required.
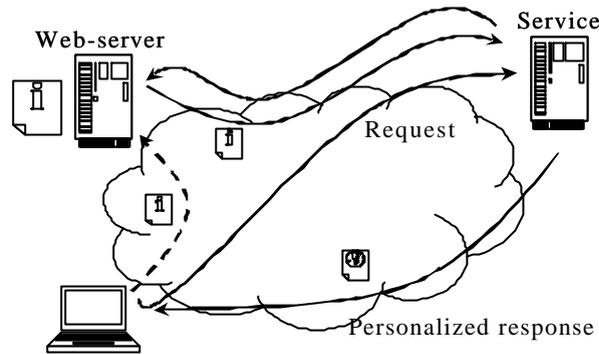

Figure 2. Third party approach

User: When the information is stored at the user end, the user can easily access and change the information. It is more likely that the user upkeeps his information locally than in some remote server. Two locations for the user-based approach have been identified, namely a personal trusted device and a personal web page.

Personal trusted device: Personal trusted device (PTD) could be anything from mobile phone to PDA. The user carries the device most of the time himself. The information is accessible by the service whenever user device is with the user. This makes it possible for user to control and update the information. If the service providers follow the idea and don't store the personal information they gain, user has pretty much the control of the security of his own personal information. This also allows the use of personal information in services that the user has not used before. Figure 3 presents the personal trusted device approach.
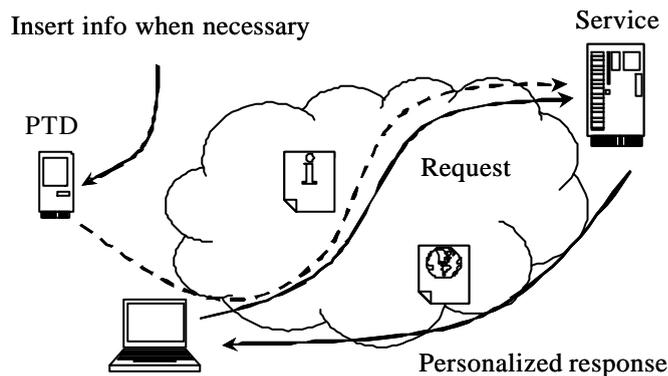

Figure 3. Personal trusted device approach

While new wireless technologies are applied to mobile devices to provide ad-hoc communications, there is no standard for exchanging various types of personal information. SyncML [SYN] supports vCard [IMC], which can be used to exchange information held in business cards, but there is no support for other type of personal information [HAN]. This might slow the operation of this approach.

Personal website: Since personal trusted devices tend to have small amounts of memory, only part of the personal information fit in it. The more space taking information can be centralized in persons own website. The data in website is controlled by the user and is accessible by the service whenever user allows it to be so.

Table 1 presents a comparison of the properties of each approach. Different storage locations have their upsides and downsides. Current mobile devices don't have very much space to store the information, compared to the dedicated servers. On the other hand, Compact flash cards, such as IBM microdrive can already hold as much as 1GB of information. As the new mobile devices support Compact flash standards the storage space on mobile device is no longer an obstacle.

Table 1. Properties of different personalization information locations

Properties of location	Service	3rd party	User: mobile	User: website
Space available	Megabytes	Megabytes	Kilobytes	Megabytes
Operation	User types in all the information for each service.	User updates 3rd party. 3rd party server sends the information to the service when required.	User updates information and takes care that information is send to the service when necessary.	User updates website and the server uses website as 3rd party.
Information update	Many places	One place	One place	One place
Accessibility for the service	Always accessible to the appropriate service.	Requires connection to 3rd party server and a contract with 3rd party.	When personal trusted device is at communication range of service accessing device.	Requires connection to the user's website
Accessibility for the user	Requires connection to the service	Requires connection to 3rd party server.	Requires the mobile device.	Requires access to the website
Form of data	Can be service specific	Requires standard	Requires standard	Requires standard
Security	Depends on all the services where information is stored.	Depends on the security of the 3rd party server.	Depends on the security of mobile device.	Depends on the security of the website.

Another big difference is the first time use of the service and the initialization of the personalization data. On service specific approach, user has to type the personalization information to every new service he uses. For other three approaches it is enough to type it only once. This also means that if information changes, it has to be updated only to one place. In order to this be true in third party approach the service provider has to have a contract with the same third party as the user for getting the information.

When using third party personalization server or personal website the personalization requires reliable connection to the given server. This means that personalization information might not be available in all places. Also unreliable network or poorly managed server can prevent the usage of information. Personal trusted device approach requires common communication medium, such as Bluetooth [BLU].

It is not enough to get access to the personalization information, but the information has to be in a form that the service understands how to use it. When the information is stored on service this is not a problem, as service provider can define the way he stores and uses the information. On other approaches the stored information is intended for use of various services and a standard way to define and transmit the information is needed.

3. PROPERTIES OF THE PERSONAL INFORMATION

As customers provide personal information for companies, it is hardly ever thoroughly thought what is really needed and what is unnecessary by the companies. If customers feel that the information is unnecessary they do not give that or give some false information. If neglecting some part of information errors may be generated. False information leads into unwanted personalization results. Better results can be achieved by carefully considering the actual needs and the properties of the required information [KAS]. The following list presents some properties for consideration:

Necessity for storage: The length of time the services need some specific personal information may vary considerably. Some information is needed by service only for a short time and some may be required for longer time periods. For example, if a web store doesn't send any advertisements through mail, then the mail address information is needed only for the delivery of ordered items and can be asked at the moment of order. On the other hand if this information is saved by the company, there is no need to request it every time. According to this study the service should store only the information it needs when customer is not using the service. While connected the information may be retrieved from the user's device.

Stability: Some information is stable, some change rarely and some frequently. For example birthday never changes and is thus stable, mail address may change rarely and the balance of a bank account quite frequently. While designing services the stability of the requested information should be considered carefully. Information that changes should be stored at the service only for the time it is needed. Frequently changing information should be stored at the user's device so that the user may update it when necessary.

Size of the information: Current personalization approaches are mainly based on textual representation of the information. As the amount of information is still quite small it does not have too much influence of the place where it is stored. In the future other forms of presentation, such as audio-visual content, could be used for the personalization. This may affect the amount of communication and thus the optimal location of the information.

Generality of the information: While a lot of information is independent from the service user still has to type this information over and over again when he wants to use new service from new service provider. This approach makes it highly impractical to provide personalized services for one-time visitors. For example a company might provide computer with Internet access to its visitors. In order to read mail and the computer need to be customized. The visitor has to type the necessary information such as the mailbox address.

Service specific information is information that is considered useful only for the given service, e.g. user ID number for some service or customer behavior. Restaurant for example may have user specific knowledge like the fact that customer likes fish. If this information is combined with the service dependent information about the types of fish dishes customer have ordered on last few times, the restaurant can offer the customer a fish meal he has not yet tried out. *User specific* information should also be on place where various services can access it, whereas the service provider is only usable place for storing service specific information.

Origin of the information: Information can also be categorized according to its origin. Some information may be generated by the service (e.g. web browsing habits and thus some preferences) and some other information deliberately asked from the user. Although automatic information gathering might work well, some users may want to have the control over their data or are at least quite suspicious about the issue. In all cases the user has to have the possibility to limit the use of the information. Information generated by the service provider may be provided to other services with the user permission. Liberty project offers an approach for service providers to share the information about the customers through the circles of trust [LIB].

Privacy: The level of privacy greatly differs according to the piece of information and the user. Users may have preferences concerning the required privacy level. Some information is very private for some users whereas the same information may be public for other users. The final decision concerning the location of the information storage has to be on the customer.

The properties of personal information affect to the optimal storage location. Based on the properties the service should carefully consider the information it needs. The final decision concerning the storage location should always be on the customer

4. PERSONAL INFORMATION DEPENDENCIES

Personal preferences are not always the same all the time, but they are dependent on the situation. When you are on a workday driving through unknown city and want to eat, you probably want to find a place where you get your lunch fast. On the other hand when you are on vacation with your wife it is more likely that you are looking fancier place to eat than a burger place. During the research five dependencies were identified.

Time: When talking about personal preferences time matters. While user might want to get latest news on his mobile phone with loud beep at the day time, at the night time such notes are more likely to be annoyance. Time is rather easy to determine by the service especially if the time zone where the user resides is known.

Location: Location of the user may affect the needs of the user. While quick burger lunch is ok at hometown, more exotic choice might be attracting, when visiting another country. Location can also affect the time parameter. At the wintertime a user might prefer hot rum, when he is in Finland. On the other hand, when he is visiting Australia at the January something cold is more likely the user's preference.

The usual way to determine rough user location on the Internet services is to ask it from the user. On services accessed with mobile devices the location information can be provided by the user device that has GPS (Global Positioning System) capability or by sophisticated service access network. [JÄP].

Role: People have many roles on their lives such as worker, father or soccer team captain. Depending on the role the preferences change. As the worker the user will more likely to select different books than as the father.

Situation: Sometimes time, role and location do not give accurate enough information about the user preferences and the user's situation is needed. Phone calls for special offers are probably not welcomed when user is at lunch or at the meeting.

The role as well as the situation in which the user is at the time is hard to predict. In some extent situation can already be determined in some mobile phones, e.g. Nokia 6110. Those have adjustable profiles function for different usage situations with call grouping and caller group identification [NOK]. Similar approach could be used also for the role and thus the mobile device could provide the information of role and situation of the user.

Equipment: Equipment affects the representation of the information. Information should be provided in a different way on a desktop or on a PDA device. This may limit the presented content [HJE].

The dependencies do not directly affect the location of the personalization information but they have to be identified when the user uses the service.

5. WAYS FOR USING PERSONAL INFORMATION

A lot of information may be gathered for the personalization purposes. These personalization attributes are then used to choose various service parameters or attributes. The following list shows four general methods how these attributes of personalization can be used to determine service attributes and to provide personalized service.

Direct: The information is used as it is and nothing is compromised, e.g. if the user prefers black background with white letters then he gets them. In this model there is no need for computing power to provide the personalization to the customer. To determine one service attribute one personalization attribute is enough.

Indirect: The information is used indirectly to provide personalized service, e.g. if the user prefers black background with white letters, service provides interface with dark background color and light letters according to the official color schema of the company. Very little computational power is needed for generating the correct service attribute. One personalization attribute can be used to determine many service attributes.

Approximated: Personalization is approximated from the various types of information, e.g. if person likes sci-fi movies and holds huge sum of spare money, he is suggested to buy a shuttle trip to the space. Approximated model requires a lot of computational power to generate the service attribute from the set of personalization attributes. Same attributes in different sets can be used to determine many service attributes. In approximated mode either the service attributes or personalization attributes can be stored to provide the personalized service when customer returns.

Profiled: User's personal information is used to profile the user and then to determine which predefined group the user belongs [ARD]. After that the user is treated as the member of this group, i.e. the actions are performed as for any person in this group. The profiles have been generated beforehand and thus little computing power is needed when determining which profile is used for service creation. Many personalization attributes are used to choose one service profile containing the attributes. After profile is created there is no need to store the personalization attributes used for generating the profile.

Table 2. Usage of personal information

	Direct	Indirect	Approximated	Profiled
Required computation when using information	None	Little	Lot	Little (creation of profiles requires a lot)
Source information: Personalization attribute	1:1	1:M	M:M	M:1
Stored information	Personalization attributes	Personalization attributes	Service or personalization attributes	Profile
Required communication	Amount of personalization data	Less than in direct model	Lot	Little from one user
Required storage space	Lot	Lot	Lot	Little

The first three ways are more or less personalization whereas the last way is more like grouping. Considering the needs of every person is much more difficult than considering the needs of only a predefined number of groups. These different ways of using personal information are compared in Table 2.

When information is used directly, the amount of information needed can be rather big as one piece of personal information determines only one personalization attribute. This also results to heavy traffic when personalization is used. In the case of indirect and approximated models the situation is almost the same. A lot of storage space is required for the personalization information. The difference is in the amount of computation required.

The profiled method differs from the others in a way that once the user is categorized into one profile it rarely is changed. Therefore, the amount of transferred data remains low. Optimally different profiles are stored at the service while the user device holds the identifier of which profile the user matches at. This ensures that once the user has profiled he don't have to send his personalization information again to the service thus reducing the amount of data transfer and providing better security as the user's identity is not linked to the profile at the service database.

From the remaining methods indirect model requires least data transfer and therefore the personal information can be stored on the personal trusted device and used from there on most services. Approximated model requires lot of information about the user to generate various personalization attributes therefore the information should be accessible to the service rather easily. Internet storages or at the service itself are recommended places.

The optimal place for personal information storage in direct model depends on the amount of personal information needed for service. When little is needed the best place is to have it at user device, but if a lot is needed the transmission medium becomes bottleneck on service usage. In this case the information should be stored at the service.

6. CONCLUSIONS

The objectives of this study were to minimize the amount of redundant personal information. This goal was achieved though the analysis of location, properties and dependencies of the personalization information. As the result of this analysis the usability of services and the security of private information will be increased.

Overall the optimal place is not just one location but the combination of given locations. Generally the user specific information should be stored at the personal trusted device. Services should store only the service specific information and the information they require when the user is not using the service. 3rd party should be used when forming circles of trust in which services are allowed to share their own specific data or exchange authentication information as is done in Liberty alliance's approach.

In order to exchange the personalization information a method for transferring the information is needed. In our study the Bluetooth wireless communication technology has been used for connecting the user's device into the network. Although this is not discussed in this paper it should be considered as a candidate for transfer media from PTD to service accessing device. Also to use the information a notation for describing the personalization information is required. Currently service providers have their own proprietary solutions.

Several independent projects have been defining XML based markup for some personal information. Unfortunately none of these are widely accepted.

Right information in right place can increase usability of the services of e-society dramatically. When user has device holding the information, the service provider needs less space for storing user preferences and user can determine a bit better where his information is going and thus decide the balance between usability and privacy.

REFERENCES

[ARD] Ardissono, L and Goy, A. 2000. Tailoring the Interaction with Users in Web Store. *User Modeling and User-Adapted Interaction*, 10(4), pp. 251-303

[BLU] Bluetooth SIG, 2002. *Bluetooth specification 1.1*. http://www.bluetooth.org/. Accessed 28.2.2003

[HAN] Hansmann, U. et al. 2002. *SyncML, Synchronizing and Managing Your Mobile Data.* Prentice Hall, Upper Saddle River, USA

[HJE] Hjelm, J. 2000. *Designing Wireless Information Services*. Wiley Computer Publishing.

[IMC] Internet Mail Consortium. *vCard and vCalendar.* http://www.imc.org/pdi/. Accessed 28.2.2003

[JÄP] Jäppinen, P. and Porras, J. 2002. Architectural Considerations on a Bluetooth based Guidance System, *Proceedings of the IASTED International Conference WIRELESS AND OPTICAL COMMUNICATIONS*, Banff, Canada, pp.281-286

[KAS] Kasanoff, B. 2001. *Making it Personal*. Perseus Publishing

[LIB] Liberty Alliance, 2002. *Liberty Architecture Overview*, Available at: http://www.projectliberty.org/. Accessed 29.5.2002

[NOK] Nokia,1999. *Nokia 6110 specification*, http://www.nokia.com/phones/6110/specs.html. Accessed 25.1.2003

[PAS] Microsoft, 2002. *Microsoft .NET Passport.* http://www.passport.com/. Accessed 29.12.2002

[SCH] Schneier, B. 2001. *Secrets and Lies: Digital security in a networked world*. Wiley Computer Publishing.

[SYN] SyncML Consortium, 2002. SyncML Specification. http://www.syncml.org. Accessed 28.2.2003

[VOT] Voth D. 2001. Personalization vs. Privacy. *CRM Magazine* August 2001

[YRJ] Yrjölä M. et al. 2003. Personal information transfer from mobile device to web page. Submitted to: *3rd Workshop on Applications and Services in Wireless Networks*. Bern, Switzerland