On Information Exchange for Virtual Identities: Survey and Proposal

Dawid Grzegorz Węckowski
Poznań University of Economics
Department of Information Systems
Email: dawid.weckowski@kie.ue.poznan.pl

Jacek Małyszko
Poznań University of Economics
Department of Information Systems
Email: jacek.malyszko@kie.ue.poznan.pl

***Abstract*—The emerging concept of User Virtual Identity on the Web is inevitably related to information exchange between different entities. Therefore, we analyze the current solutions of an Identity-related information exchange, taking into consideration categories of the information being exchanged, parties involved in the exchange process as well as the exchange protocols. The analysis allows us to define an information exchange solution for the project Ego – Virtual Identity.**

***Keywords*-virtual identity; information exchange;**

I. INTRODUCTION

Ever increasing amount of information is exchanged every day on the Internet. Users reveal bits of their interests and preferences while surfing the Web and using e-services. Building a Virtual Identity, which can comprehensively deal with the user representation in that environment, is inherently bound with information exchange [1].

We argue, that there is a need for a solution that will assure user-centricity in the field of Virtual Identity information management at the same time providing a level of anonymity. In this article we present an analysis of the state-of-the-art Identity-related solutions with the aim of finding the strengths and weaknesses of their information exchange implementations. Than we propose an information exchange model for the Ego – Virtual Identity system, that will preserve user anonymity, still enabling robust personalization, which potentially would allow building of a central Identity information store supplied with information from various sources. With this model the users can achieve a level of anonymity and tracking protection that can't be achieved in current solutions.

The article is structured in the following way. In Section II we consider personal information categories and accompanying solutions. Section III comprises analyzes of parties involved in the information exchange process. Details of the processes in different solutions are presented in Section IV. Then we present the information exchange model for the Ego project in Section V. The article concludes with the final remarks.

II. INFORMATION BEING EXCHANGED BY VIRTUAL IDENTITY

Type and amount of information being shared and acquired by a user's Virtual Identity is supposed to depend on the user's information needs and preferences regarding sharing this information with services. That brings us to a conclusion, that it is the user who is supposed to decide which information is to be exchanged. The user is the ultimate source of knowledge about the value of the information based on his or her needs and preferences.

A. Personal Information Sources

Personal information can be perceived as an information regarding a person and created by the person [2]. There are several types of personal information depending on the sources, from which such information is being acquired:

- *Volunteered information* — an information that is shared by a person freely and explicitly, usually by providing descriptive resources about oneself (e.g. filling a form with one's interests, publishing a CV) or personally generated content (documents, music etc.)
- *Observed information* — an information that is gathered by recording person's activities, while he interacts with various devices and applications (e.g. capturing Internet-browsing history, GPS location). This information is collected implicitly, with no additional user's actions.
- *Inferred information* — an information that is a result of reasoning process based on other personal information. It is often being performed by institutional bodies for widening the knowledge about a user, eg. client's financial history can be used to calculate credit scores by a bank.

B. Personal Information Categories

In the literature there are many attempts to enumerate types of personal information. The most popular approach is to categorize the information according to functional perspective, as it may be found in FIDIS deliverables [3], GUMO ontology [4] or in the Marc Davis' talk for the World Economic Forum [2], [5]. More user-oriented approach can be found in the work of Brusilovsky [6]. Additionally, Mitchel et al. [1] consider types of volunteered information being shared between enterprises and individuals.

We believe that, apart from personal information categories described by Nabeth [7], there is a more important perspective that describes a user — the *user's perspective*.

 ISBN: 978-1-61208-249-3

This perspective categorizes the information with respect to four essential characteristics of the users – who they *are*, what they *know*, what they *have* and what they *do* – as it was introduced by Anrig et al. [3].

Any superposition of pair of those characteristics can be perceived as a category of personal information. Thus the users can be characterized by the following types of information:

- *Attributes (are + has)* – any user's feature that can be directly described by an observer or can be extracted from any institutional records, including demographic data or biological features.
- *Acquisitions (have + know)* – all knowledge and possessions, including physical and virtual goods, both generated and consumed.
- *Roles (are + do)* – any relations that describe the user, inclusive of profession, citizenship or social affiliation and roles.
- *Abilities (know + do)* – any user's competences related to tacit knowledge the user has, as well as any user's activities that are indicators of those or any other user's features.

Additionally, we would like to propose two other categories as follows:

- *Context (do + have)* – information about user's relations to the external world and any objects of that relations, including people, locations or events.
- *Self (are + know)* – meaning reflective consciousness, any user's features that cannot be observed directly but rather can be deducted or frankly expressed by the user, as features of that category are strongly related to the user's state of mind, such as personality, preferences or interests.

The categories and corresponding essential characteristics are shown in the Figure 1.

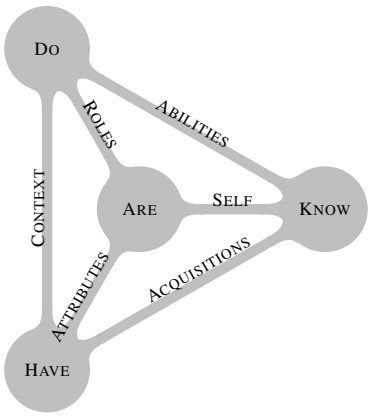

Figure 1. Personal information categories from the user's perspective

III. PARTIES INVOLVED IN THE PROCESS OF INFORMATION EXCHANGE WITH THE IDENTITY

In this section, we will analyze different entities, that exchange the information with the Identity. Such exchange can be carried out in two directions; there can be an outbound flow of information (from the Identity to the outside entities) as well as an inbound transfer (external entities send information about the user to the Identity).

A. The consumers of information

The most important class of entites, that is mentioned in projects concerning the consumers of information stored in the user Identity, are Service Providers. In many situations, services need to learn attributes (permanent or temporary) characterizing their users.

The "Service Providers" term covers a wide range of entities with many different purposes for which they need the information. Probably, the most important of such purposes is authentication and authorization of users. For example, with prototype created in one of the STORK project's pilot programs, students can use foreign university services, proving their real Identity with eID cards, issued by their domestic government Identity providers [8].

Currently, there are more and more services, which need Identity information to fulfill other goals. Adaptive systems personalize their content and other features to users' needs. The simplest example are personalized Web portals, for example news sites. This domain is traditionally in interest of user modelling; based on user attributes representing her interest, the contents of a news site is adapted to match user's needs, as a result generating a personalized online magazine [9], [10]. Here, the consumer of information from an Identity is the provider of this web portal [11].

E-commerce services may use a so-called recommender systems, which, based on a user preferences, inform the user about products of the provider, which may best suit the user's needs [12]. These parties are consumers of Identity information also in ProjectVRM [13], where users can publish their needs (intents to buy specific items) on a so-called Personal Request for Proposals platform, and any vendors using the platform can thus learn users needs and respond with a personalized offer [1].

Mobile and location-based services can utilize user's location and attrbiutes in many ways, for example alerting the user when one of his friends is nearby [14].

Technically, it is possible to broaden described types of services to more real-world entities [15], [16]. For example in Ambient Intelligence vision, a temperature of water in the shower in the hotel room can be automatically adjusted to preferences of the roomguest, or music played in the restaurant may depend on average preferences of its guests. In these examples, service providers consuming the Identity information would be devices in such places as restaurants or hotels.

 ISBN: 978-1-61208-249-3

Another type of service providers are public institutions, such as hospitals. In scenario described in [16], a person's up-to-the-hour health information collected by a medical device attached to the person's wrist can be transferred to her hospital. Similarly, an information about user's location can be sent to the hospital during an emergency call [14].

The information about the users may be also transferred between two Identity Providers in order to merge user accounts or to reconcile differences between such identities. This is the case in Federated Identity scenarios [17].

Finally, consumers of Identity information are also user devices. The users may want to see their Identity information and in such situation it must be transmitted to his device. The information may be also transmitted to other users' devices and displayed there, if its owner agrees.

B. Information exchange – the inbound flow

The goal of inbound flow of information exchange is to update the Identity with additional information about its owner. Entities, that are sources of such information, can be the users (when they manually input some information to the Identity), a special sofware aiming to track the users' behavior to learn their attributes, or, theoretically, even Providers of services.

Editing information stored in the user model by it's owner is called *explicit user modelling* and was a traditional way of building user models in many adaptive systems in the past [18]. In this case, the information is transmitted from a user's device to the Identity.

The other method of feeding models with information about the users is *implicit user modelling*, in which users' behavior is analyzed (with or without their awareness and consent) in order to learn their attributes, without users' involvement [18]. This can be achieved in many different ways: by monitoring web log files, tracking pages visited by the users or queries their input to a search engine, etc. [9], [19]. An interesting example here is a mechanism of scrobblers used by Last.fm [20] portal, which maintains user music profiles. Information is sent from many such scrobblers installed on multiple devices used by the owner of Last.fm account to a single profile, stored on Last.fm servers.We believe, that such tracking mechanism can be extended to other user activities as well.

IV. Information exchange process

A. Front channel / Back channel

Identity-related information exchange involves passing some assertions about a user from an Identity Provider to a Service Provider and *vice-versa*. Most common approaches to that process are [21]:

- *Front channel* – information being exchanged by a redirection of a user's browser from one site to another with custom defined parameters. User can be informed in details about what is being exchanged.
- *Back channel* – direct information exchange between an Identity Provider and a Service Provider, after previously establishing an association, without a user participation. The user has no knowledge on what is being exchanged.

In subsequent sections we will analyze existing solutions in the implementation of Identity-related information exchange process. We will focus on the most popular projects.

B. Authentication-oriented solutions

Currently, two main authentication-oriented Identity solutions can be mentioned: popular and well-established OpenID [22] , and WebID [23], which is currently being developed.

The communication process, using authentication-oriented solutions, can be summarized as follows [22], [24]:

1) A user requests resources from a Service Provider and is asked to authenticate.
2) The user supplies the Service Provider with an identifier of a Virtual Identity.
3) The Service Provider verifies, if the user is the owner of the claimed Virtual Identity, eg.:
   - the OpenID redirects the user to the Identity Provider's website and verifies the response assertion, stating if the user has managed to login successfully,
   - the WebID compares user-provided certificate with the user public key published on the Virtual Identity website.
4) If the authentication succeeds, the user is provided with the requested resource.

C. Authorization-oriented solutions

Another group of the Identity-related solutions are those which allow for exchanging complex information for authorization purposes, eg. SAML and OAuth.

SAML [25]is an OASIS standard, defining a framework for describing and exchanging security information with the use of XML. While SAML can be used in different business scenarios, our interest is mainly focused on establishing Federated Identities and Multi-Domain Single Sign-On.

Another solution, OAuth, is a protocol allowing users to authorize third parties (here called clients or consumers) to access server resources, owned by the users, without revealing their credentials to the clients [26], [27].

D. The Identity Metasystem

The Identity Metasystem is based on Kim Cameron's identity laws and uses Information Cards for representation of Digital Identity [21], [28]. The exchange model is based on SOAP messages and uses a number of OASIS standards for Web services (WS-Trust, WS-SecurityPolicy and WS-MetadataExchange) [29].

 ISBN: 978-1-61208-249-3

The central elements of the metasystem from a user's point of view are Information Cards, that are visual representations of different digital Identities of the user. Such cards are presented to the user by a software component called Identity Selector, which allows users to easily choose from available cards [21].

Also, many other, smaller Identity-related projects exist. Example lists of such solutions can be found at: [30], [31]. Having analyzed the most popular solutions we developed an information exchange model for the purpose of the Ego project.

V. Information exchange model for the Ego Project

A. The goal of the Ego information exchange model

The Ego project aims to research the possibilities of improvements in the area of instant personalization based on exchange of an Identity information during everyday tasks performed by the user during browsing the Web (for example visiting different portals in search for interesting articles or items). At the same time, we want to ensure, that the users can achieve a complete anonymity and complete control in terms of information, that service providers have about them.

Much research has been conducted on different aspects of users' privacy, anonymity, Identity management and global user modelling. Still, we have noticed a serious drawback in the current solutions. The Identity can store a lot of information about the user, but at the same time service providers store a lot of user-related information on their side, building user models for their own needs. Such information is treated by service providers as an important asset, giving them a competetive advantage on the market. This fact has two negative implications:

- users cannot easily learn, what Service Providers know about them, or update such information, if it is not accurate,
- users cannot reuse this information in another services.

In the following sections, we will describe our position on how this situation can be changed.

B. User session at Service Provider's website

The main goal, that we want to achieve with our information exchange model, is that the information used for personalization should be stored on the Identity Provider's side and not on Service Provider's servers. Thanks to that, users would be able to easily see and change any information, that different service providers may have about them.

To achieve that, the exchange model must be extended to foster an exchange of information in both ways (inbound and outbound). Service Providers must be encouraged or enforced to send information, that they have about the user, to the user's Identity. We came to a conclusion, that they would do that only in one situation: when using information about the user gathered during a single session and stored on their side would be impossible in next sessions. To achieve that, Service Providers must not be able to link different sessions of the same user; such situation is called unlinkability [32].

We decided, that in our exchange scenario, the user will not provide the Service Provider with his identitfier, but only with URL to his Identity Provider. This can be thought as a next logical step in assigning different identifiers to a user based on a context, in which he is working (a so-called unidirectional Identity) [28], [32].

With such an exchange model, the Service Provider would have to send the information about the user, that he has gathered, back to the Identity. At this point, a question arises on what and how the Service Provider can store in the Identity. It must be restricted to some degree, otherwise the Service Provider would be able to put a user-specific identifier in the Identity (similarly as it is done in cookie files), and based on that to use profiles stored on his side, without sending it to the Identity. Some solutions that are possible here are:

- the storage can have a specified structure, which would restrict possible user characteristics to some predefined ones;
- user characteristics in the storage can be managed not by the Service Providers directly, but set by some Identity Provider modelling algorithms, which assign appropriate values based on user characteristics sent from the Service Provider,
- values sent to the Service Providers as responses to queries can be slightly changed. Such changes may be random or may be generated based on predictions of future user needs, at the same time helping users in finding new interesting items, similarly as in our paper [33].

C. Steps in information exchange model proposed

In this section, we will discuss the most important steps in our exchange model in greater detail.

1) Establishing a common user identifier between Identity Provider and Service Provider
   This is the first and basic step in our information exchange model. To enable exchange of information between the Identity and Service Providers about the user, an identifier must be assigned to the user. Such identifier must be randomly generated, if we want to force Service Providers to send information about the users to the Identity. This process is shown in the Figure 2.
2) Authentication of Service Provider request
   The identifier is now established and callable. It can be exposed in a form of URL, to which the Service Provider can send messages. Still, the identifier is dedicated to only one Service Provider. It must be there-

 ISBN: 978-1-61208-249-3

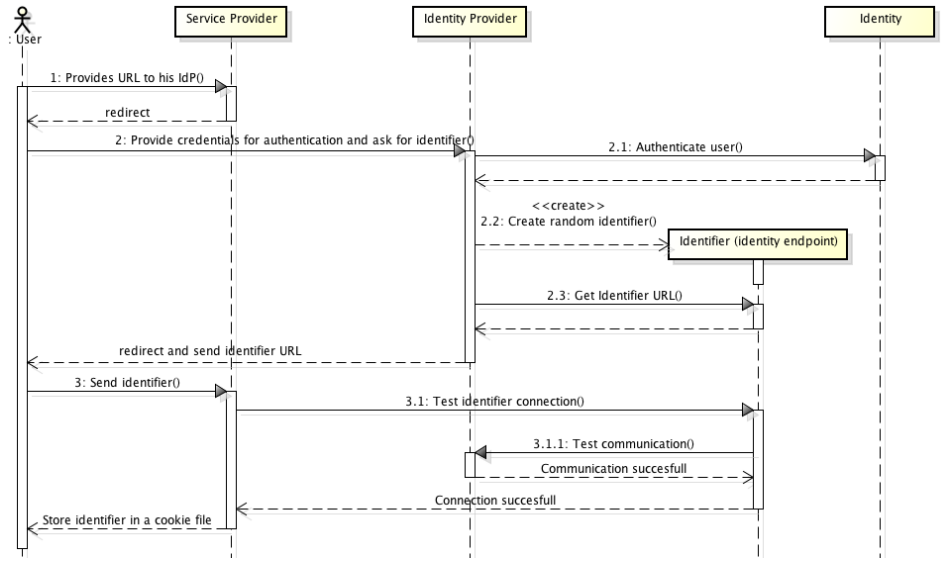

Figure 2. The process of information exchange to establish an identifier

fore ensured, that the messages, that are received at the identifier URL, are indeed sent from the intended Service Provider. A Service Provider authentication mechanism must be therefore established.

3) Service Provider's request for information from the Identity
   One of two types of message exchange requests is a query for Identity information. The Service Provider needs this information to personalize its service. The requests are sent by the Service Provider via back channel when the user wants to access a certain resource and is received by Identity Provider at identifier URL. At this step it must be ensured, that the Service Provider is authorized to use the identifier and that the transmission is secure.
4) Request for an update of the Identity by the Service Provider
   When the Service Provider has collected a new information about the user, the information may be sent to the Identity. This is a similar flow to one specified in previous step, but in the other direction. It's important to note, that the Service Provider must send it before the identifier expires.
5) Deletion of the identifier
   As the identifier is tepmporary, at some point in time it must be deleted. After that, the Service Provider cannot get more information about the user or update the Identity with new information. Such deletion can be triggered by different events, such as:

   - user can request deletetion of a specified identifier or all identifiers at Identity Provider's website;
   - the identifier can be set for automatic expiration trigger when a certain time passes since it was created.

   User should also have an opportunity to configure a certain identifier to be persistent, so that it would not be deleted. Users may chose to do so for example, if they want to have a long-term and closer relationship with a certain Service Provider.

VI. Conclusions and Future Work

In this paper, we have analyzed different issues related to exchange of user's personal information on the Web. The most important existing solutions were presented in terms of conceptual and technical means that they use to exchange the Identity-related information. Based on that, we have proposed our own information exchange model, which enables users to gain better control on information, that different entities may have about them. Our solution ensures, that all information, that Service Providers have about the user, will be stored in the user's identity and under the user's control.

Our further research directions focus on development of a reusable user model structure, which must be universal enough to be usable by many different adaptive systems, for example tech news portal, music recommender systems etc. Apart from that, we plan to develop a mechanism that would enable updating this model based on information sent from Service Providers. A comprehensive policies mechanism is also planned, that would allow users to clearly define conditions, under which their personal data may be exchanged with third parties.

 ISBN: 978-1-61208-249-3

ACKNOWLEDGMENT

The work published in this article was supported by the project titled: "Ego – Virtual Identity" (http://kie.ue.poznan.pl/en/project/ego-virtual-identity), financed by the Polish National Centre of Research and Development (NCBiR), contract no. NR11-0037-10.

REFERENCES

[1] J. Andrieu and J. Clark, "The information sharing report," Kantara Initiative, Tech. Rep., 2010.

[2] World Economic Forum, "Personal Data: The Emergence of a New Asset Class," http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf, January 2011, accessed: 11-02-2013.

[3] B. Anrig, E. Benoist, and D.-O. Jaquet-Chiffelle, "Virtual? identity," in *FIDIS Deliverable 2.2: Set of use cases and scenarios*, T. Nabeth, Ed. FIDIS, 2005, pp. 22–34.

[4] D. Heckmann, T. Schwartz, B. Brandherm, M. Schmitz, and M. von Wilamowitz-Moellendorff, "Gumo — the general user model ontology," *User Modeling 2005*, pp. 428–432, 2005.

[5] K. Hamlin and M. Hodder, "PDEC response to FTC DNT White Paper," http://1.usa.gov/pdeDNT, February 18 2011, accessed: 11-02-2013.

[6] P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive educational systems," in *The adaptive web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 3–53.

[7] T. Nabeth, "D2.3: Models," FIDIS, Tech. Rep., October 2005.

[8] D. Berbecaru, A. Lioy, M. Mezzalama, G. Santiano, E. Venuto, and M. Oreglia, "Federating e-identities across europe, or how to build cross-border e-services," 2011.

[9] W. Abramowicz, *Filtrowanie informacji*. Poznań: Wydawnictwo Akademii Ekonomicznej w Poznaniu, 2008.

[10] J.-w. Ahn, P. Brusilovsky, J. Grady, D. He, and S. Y. Syn, "Open user profiles for adaptive news systems: help or harm?" in *WWW '07*. New York, NY, USA: ACM, 2007, pp. 11–20.

[11] M. Koch, *Global Identity Management to Boost Personalization*, 2002, pp. 137–147.

[12] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[13] http://cyber.law.harvard.edu/projectvrm/, accessed: 11-02-2013.

[14] A. Deuker, "D11.2: Mobility and lbs," FIDIS, Tech. Rep., July 2008.

[15] W. Schreurs, M. Hildebrandt, M. Gasson, and K. Warwick, "D7.3: Report on actual and possible profiling techniques in the field of ambient intelligence," FIDIS, Tech. Rep., August 2005.

[16] P. Scholta, "Refined swift scenarios, use cases and business models," *SWIFT Project Deliverable*, 2010.

[17] N. Ragouzis, J. Hughes, R. Philpott, and E. Maler, "Security Assertion Markup Language (SAML) V2.0 Technical Overview," http://www.oasis-open.org/committees/documents.php?wg_abbrev=security, 2006, accessed: 11-02-2013.

[18] A. Kobsa, "User modeling: Recent work, prospects and hazards," *HUMAN FACTORS IN INFORMATION TECHNOLOGY*, vol. 10, pp. 111–111, 1993.

[19] M. Hildebrandt and J. Backhouse, "D7.2: Descriptive analysis and inventory of profiling practices," FIDIS, Tech. Rep., June 2005.

[20] http://www.last.fm/, accessed: 11-02-2013.

[21] C. Burton, "The Information Card Ecosystem: The Foundamental Leap from Cookies & Passwords to Cards & Selectors," http://www.w3.org/2005/Incubator/webid/spec/, 2011, accessed: 11-02-2013.

[22] O. Community, "OpenID Authentication 2.0 - Final," http://openid.net/specs/openid-authentication-2_0.html, 2007, accessed: 11-02-2013.

[23] http://www.w3.org/wiki/WebID, accessed: 11-02-2013.

[24] H. Story and S. Corlosquet, "Web 1.0. Web Identification and Discovery," http://www.w3.org/2005/Incubator/webid/spec/, 2011, accessed: 11-02-2013.

[25] http://saml.xml.org/, accessed: 11-02-2013.

[26] E. Hammer-Lahav (Ed.), "The OAuth 1.0 Protocol," http://tools.ietf.org/html/rfc5849, 2011, accessed: 11-02-2013.

[27] J. Camenisch, J. Riordan, and S. Sandra, "D 1.1.1: Analysis of existing web protocols for trusted contents," PrimeLife project deliverable, Tech. Rep., August 2008.

[28] K. Cameron, "The laws of identity," http://msdn.microsoft.com/en-us/library/ms996456.aspx, 2005, accessed: 11-02-2013.

[29] M. Jones and M. McIntosh, "Identity metasystem interoperability version 1.0 (imi 1.0)," *OASIS Standard*, 2009.

[30] http://wiki.idcommons.net/Identity_Landscape, accessed: 11-02-2013.

[31] http://personaldataecosystem.org/2011/06/startup/, accessed: 11-02-2013.

[32] A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management," *URL: http://dud. inf. tu-dresden. de/literatur/Anon_Terminology_v0*, vol. 34, 2010, accessed: 11-02-2013.

[33] W. Abramowicz, J. Małyszko, and D. G. Węckowski, "Discovering of users' interests evolution patterns for learning goals recommendation," vol. 90, pp. 231–238, 2011.

 ISBN: 978-1-61208-249-3