DIGITAL IDENTITY: HOW TO BE SOMEONE ON THE NET

Toby Baier, Christian Zirpins, Winfried Lamersdorf
Universität Hamburg, FB Informatik, Verteile Systeme und Informationssysteme
Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
baier|zirpins|lamersd@informatik.uni-hamburg.de

ABSTRACT

Personal communication and collaboration has been and still is a major driver of the Internet. A severe drawback in human centric electronic interaction is the fuzziness of the image that the co-operation partners have of each other (i.e. their respective "identities") – especially in different and varying application contexts. This uncertainty adversely affects increasingly important "soft" co-operation factors like, e.g., trust and social behavior, and should therefore be minimized whenever possible. In addition, the lack of a homogenous representation of digital identities results, even at the system-level, in many cases in increased and unnecessary administration tasks – like, e.g., keeping track on user-ids and passwords or typing the same information several times. This makes communication inefficient and error-prone and may introduce various privacy threats. On the other hand, neither the minimal identity representation which is already used at the system's level (e.g. a user-id used for security reasons), nor the emerging proprietary efforts for identifying users uniquely at the application level (e.g. for "single sign on" purposes) suffice for comprising the user's identity fully as needed for co-operation of individual human beings.

In order to cope with such problems of proper electronic user "identification", we propose an open and generic notion of a *digital identity* that is generally applicable and includes an extensible set of identity facets on the system- as well as the user-level. Such a unique digital identity for all possible Internet communication and co-operation tasks enables users to recognize distinct co-operation partners uniquely in many different contexts – but also allows for revealing individual (i.e. only partial) views on such information whenever necessary. Therefore, such a facility enriches communication by semantic information about co-operation partners and thus enables faster, more secure and trustworthy collaboration.

In summary, this paper proposes the concept of a *digital identity* and specifies what challenges are to be met when building an open, distributed, decentralized *system infrastructure for digital identities*.

KEYWORDS

Personal Communication, Distributed co-operation, Digital Identity, Virtual Communities

1. INTRODUCTION

Twenty years ago, the Internet started out as a collection of technical infrastructure mechanisms that enabled data interchange between networked machines. As a fundamental aspect, the technological simplicity of TCP/IP as well as the basic infrastructure that was granted by institutional organizations resulted in a healthy size from the start and fostered immense growth. But most important, standardized means of personal communication (e.g. email and Web) that were both uncomplicated and useful, introduced a real world appliance from the very beginning that soon turned out as what marketing people call "killer application". It is important to note that personal communication is a major factor for Internet growth.

Together with the expansion of the net, its utilization went forward. The initial means of personal communication were applied to various domains like Electronic-Commerce, -Government, -Learning, -Marketing and -Publishing just to name a few. Moreover, personal communication evolved. The replacement of "snail" by "electronic" mail and publishing of brochures by Web pages was just the beginning. Meanwhile we are used to electronic forums, instant messaging and personalized Portals. And the end isn't reached yet: we are on the way to advanced interaction concepts like virtual communities and also virtual reality (VR).

However, in all the development there hasn't been much evolution of infrastructure to directly support application-level- communication. The net stopped at the border between technical communication and its applications. For example, the only common denominator for the network representation of users in personal

communication is given by email addresses – unsuitable because insignificant and ambiguous. Anything else has to be taken care by specific applications.

As a basic requirement, most applications depend on a representation of users that enable authentication or authorization of some kind, and most of them deliver an own implementation of access control. While there have always been attempts to realize a single sign-on mechanism for various fields (e.g. Kerberos, Plan9, Microsoft .NET passport or Liberty Alliance), none of them had a major breakthrough in terms of multi-application, multi-platform, multi-media or multimodal usage. Furthermore, users often want to share additional information about their identities. For example, users of an online store might want to expose their preferences for products and users of multiple web forums might want to reveal facts of their curriculum or association with other virtual communities. This information is sometimes stored in directories (e.g. LDAP), but hardly usable in a homogenous way all over the net.

Considering personal communication and an image of the other side, one could speak of *Identities* (written with upper-case "I" to separate it from other meanings, described in chapter 3). An Identity in this context not only represents a uniquely identifiable person, but also additional information characterizing it. "Character" or "personality" are other words with a similar meaning. This can include contact- and curricular-information, but also application or services related configuration or profiles and even transient information like the current interest or task. Another part of Identity is given by social affiliations to certain groups, which is also called a *social identity*. This could include sport interests ("fan clubs"), religious beliefs or adherence to philosophical positions.

There is currently neither a standard which solves the problems of authentication and authorization, nor identity information retrieval/supply (Identity management), let alone both, for the entireness of the Internet. As indicated above, both are wanted and needed, though. In this paper we argue that authentication/authorization should be combined with repositories for personal information to what we will call a *Digital Identity Infrastructure (DII)*. It is the aim of the research project "onefC" (*open network environment for Citizens*) to realize this.

The remaining parts of the document are structured as follows. In chapter two, an example scenario is given that shows how an ordinary situation of todays net-life translates into our proposed infrastructure. Chapter three outlines the main problems that have to be tackled and works out requirements for their solution. Chapter four contains a discussion of related work. Finally, chapter five closes with a summary and conclusion.

2. EXAMPLE SCENARIO

In order to outline a user centric scenario that includes several aspects of personal communication, virtual communities and electronic commerce we use the following application example:

Consider a chess enthusiast – let's name her Elli – using the Internet to follow her passion. In order to play chess, she is using a dedicated service which provides a chess client as well as a server where other players can be met and challenged. The server computes an ELO rating[1] from games against rated players. The ELO is visible for others too and also used to mutually compute new ratings when Elli plays against them. As this is a valuable service, she has to pay for it. Furthermore, Elli also participates in a web-based discussion board about chess that is free of charge. It's an open forum where, once you applied, you can post articles, share chess transcripts and read or comment articles of others. Finally, she likes to buy books and software about chess at various online stores.

Given today's technology, this usually works as follows: First, Elli has to apply for the chess-service via email, including her credit card information. Eventually, the provider grants her access to software and server via username and password. Next, Elli has to get another username and another password for the blackboard system. While she can try to get the same username, this it is not likely to succeed. Even if it works, she is well advised not to reuse the password because of security risks. Anyway, the virtual community in the forum can't be sure that this is the same Elli they might have met on the chess server. Finally, a set of username/password combinations originate from the online stores. Initially, Elli is anonymous for each store and all contact- as well as payment-information need to be entered again. Certainly, the stores have no

[1] ELO: A method of rating chess players named after its inventor, Arpad Elo

information about her chess passion, which could result in a specialized personal offering of books or software.

Now let's imagine that a general Identity Infrastructure is in place which is used by Elli, chess-service, blackboard and stores. In this case, Elli possesses a digital identity that she can use to identify herself at all services. While authentication could be done differently for each service, identification can be done via Ellies identity in all cases. This identity also contains credit card information and Elli can grant access to this information to anyone she likes – that is, chess-service and stores. The blackboard system would not gain access to this information, though. The ELO rating, computed by the chess server, is also transferred into Ellies Identity. It's signed by the chess-server, so it's assured that she doesn't change it herself.

Moreover, the Digital Identity has social effects too: Elli is recognized as an individual personality all over the network! Let's assume Elli is shy and doesn't want everyone to know her ELO rating, so when she enters the blackboard system, other users can see that she is a chess player, but not her ELO rating. Other subscribers of the chess server though will be able to associate the ELO rating they know from playing against her on the server (or just seeing her playing there) with the postings she makes on the blackboard. They will be able to utilize this recognition to personalize the forum, maybe sort the entries not by date but by "trust" in committers which was not only gained on this board. But not only in virtual communities but also in electronic commerce digital Identity can act as a vehicle to carry trust. For example, a couple of successful commercial transactions with various stores documented in the identity could result in Elli, being treated as a first class customer with discount and premium service by the next store she enters.

This example shows that Elli surely wants to be private in certain parts of the net, or with certain parts of her Identity. But in other situations, she wants to share information about herself. She wants to make herself visible to other chess players. Consider she finds a new web forum concerning chess, would she want to introduce her to the same people she met before? Consider an instant community that is initiated by a certain chess event, wouldn't she like to be known there at once, without entering information over and over?

3. MAIN CHALLENGES AND REQUIREMENTS

The main question that naturally emerges when exploring digital identity is certainly about the notion of identity itself. Research in philosophical, psychological or social science bring some clues, some of which can help to understand how a DII should be built. In mathematics, identity is a relation where each element is only with itself. So x is identical with exactly this x and nothing else, even if it equals something in value. In more common terms, it could be said that identity is the means that distinguishes one from anything other. So if you ask someone what constitutes an identity, he is likely to list a number of attributes by which one can be distinguished from others: name, date of birth, place of birth, parents. Obviously, the word identity can be used in many different meanings. The formal, mathematical meaning is opposed by the more social meaning as personality. A personality is a (logical) identity together with a set of characterizing attributes. In this text we will write Identity with a capital "I" if we use it in the sense of personality, to distinguish it from the logical meaning.

In order to find how an Identity is further structured, one has to consider the requirements of what we want to accomplish. The example scenario comprises identification, authentication and authorization as well as sharing personal information. In fact, sharing information about ones Identity on the net is a process of self-portrayal. Images of Identities can never be complete, and in most cases certain parts of the Identity shall be hidden, generating different images. Linking these images to a more complete view on the Identity is a process of recognition, but also a major privacy threat.

Identification is the method of recognizing someone as a specific individual, maybe associating him or her with an image one already holds. Authentication is the process of making sure that the identification is valid. It gets obvious that identification is the base for anything else: Neither authentication nor the other applications can go without it. While identification is apparently straightforward (e.g. universally unique identifiers), authentication can be done in many different ways. A general *DII* should enable users to choose between different authentication methods.

Further on, more than identification and authentication is needed since we want to share *general* information about an identity. The requirement to represent personal information like address, credit card number or size of shoe in one uniform format leads to the second main problem: semantics. On the one hand predefined schemas of what can be represented by an identity are clearly not open enough. The ability to

represent everything that could be required by future net-applications is vital. On the other hand it has to be assured that Identities are universally "understood", that is, semantics are preserved. But not only semantics of attributes is a problem, metadata has to be considered too. Information associated with an Identity can originate from its owner but also be added by someone else (as seen by the ELO rating in the example). So the origin of data makes a difference. It has to be assured that the source of any information can be located and proved.

Ensuring security and privacy is crucial like in all distributed communication systems. Generally, it's imperative that disclosure of identity is fully controllable by its owner in a fine granulated way down to single pieces of information. An Identity contains a variety of information about users that they certainly don't want every communication partner to see at once. Elli would not want the blackboard system to see the credit card number she entered into her Identity but the online stores are granted access. Subsequently, there needs to be the possibility to show certain information to some communication partners and hide it from others. The more this differentiation can be mended, the better privacy can be assured. It has already been shown that sophisticated Identity management can have positive impact on privacy [Berthold, O. and Köhntopp, M. 2001]. The problem is about doing this securely. Deception and Identity theft are major threats that could lead to severe social problems. It has to be assured that Identities and even certain attributes cannot be stolen or counterfeited. Moreover for most communication acts it must be certain who is communicated with, even if this is often application dependant. If there is no trust in system security, communication partners are not trusted either.

Fine granulated information disclosure bears another problem, though: the management of Identity and access control might become a very complicated task, even with sophisticated tools. Authorization of information access cannot always be done by direct interaction. The use of rules or policies can become hard to handle too. An integration of general privacy systems like the Platform for Privacy Preferences (P3P) project [Cranor, L., Langheinrich, M. et al 2002] could make this task more flexible.

Nevertheless, too much automation in privacy control leads to another problem: the more information is exchanged automatically, the less the user is *aware* of what others know about her. The term awareness in computer science is most commonly attached to *presence awareness*, for example in CSCW or instant messaging applications. We will use the term in a more common sense: awareness is the consciousness of what is happening; in this case awareness is about the consciousness of the communication partner's identity as well as how much information is given out. Today, most Internet users are not aware of what information about them can be seen. Using the internet generates traces, but these traces are rarely visible to the user. For example, the search engine Google (www.google.com) keeps track of every search term which is entered, and also the IP address the search was originated from. The awareness problem has to be reflected in computer-human-interaction (CHI) of communication- and administration applications, taking into account questions of usability and ergonomics. Especially CHI aspects of Identity management applications that control information associations and the access, is a very important matter for awareness and therefore for privacy. By improving awareness digital identities contribute to a major problem of privacy. A general *DII* together with ergonomic software should aim at detailed control of information spread (e.g. traces). Logging and analysis of information access and communication is an important matter here.

Another aspect of Identities is not directly shown by the example scenario, although it could be indirectly derived. Sometimes people do not just identify themselves as unique individuals, but rather as part of a group. While this seems to be just another label, it is rather a true part of personality, described in detail by the concept of "social identities" [McGarty, C., Haslam, S. A., Hutchinson, K. J. & Turner, 1994]. In the example, Elli could feel herself belonging to certain social groups, namely that of chess players or that of the blackboard community. While digital representations of social identities ease assigning oneself to a certain group, they also ensure a certain level of anonymity. This can be done by uttering affiliation to a social identity while hiding personal information.

A more general problem which applies to all collaboration systems is the need for a broad user base. In the example scenario it would be of little use for Elli to maintain her Identity if she cannot use it in all the places she visits on the net. The more places one can use an identity, the more useful it is. And the more users a system has, the more useful it gets too, because more peers can be identified. Therefore, a general approach to DII has to foster its wide adoption by maximum openness, effectiveness and efficiency.

Finally general requirements for successful distributed systems apply that have been extensively discussed and will only be sketched here. Security has been mentioned already and is very important. Reliability of service is also very important. For reliability the system should be kept simple enough so it can

be administered easily and doesn't break down from maintanance overhead. Decentralization also is a crucial point for an Identity infrastructure. A centralized system that holds all Identity information of all users in one place would not only be a single point of failure and a bottleneck, but also a non-acceptable security risk.

4. RELATED WORK

The open network identity infrastructure we are proposing revolves around issues of security mechanisms and data repositories. In our approach to support open systems for personal communication, we target most directly on global user authentication and open distributed directory services. In this section we will outline related work in those areas and compare it to our approach.

Since multi user capabilities were introduced to computing systems, users had to authenticate themselves before being granted access. This became more complicated when computers were interconnected via networks. There have always been efforts to make authentication easy and transparent to the user, but secure and reliable for the system. However, most of the mechanisms which evolved, like Kerberos [Kohl, J. and Neuman, B. C., 1993] or more recently Plan9 [Cox R., Grosse, E., et al, 2002], were bound to one computing architecture though and therefore not applicable to the entire Internet. Recent development shows that two main players compete for single sign-on dominance on the Web: Microsoft's ".NET Passport" and the Liberty Alliance led by SUN Microsystems.

The .NET Passport System by Microsoft [Microsoft Corporation, 2002] aims to achieve single sign-on for web sites using only standard protocols and no extra software components on the client side. When a user tries to log into a passport enabled web site, she is redirected to the .NET Passport Server (www.passport.com) on an encrypted connection (HTTPS) and asked to authenticate via email address and password. If succeeded, the Passport server redirects the user back to the original site and sets a cookie on the client host that the server can read to see who is logged in. This architecture has several flaws and has been hacked or broken several times already [Electronic Privacy Information Center, 2002].

The Liberty Alliance Project [Liberty Alliance, 2002] aims to establish another single sign-on infrastructure for web applications. The proposal focuses on federations of web services and applications so, once granted, they can share user information, including authentication. Like in Passport, a change of client applications is avoided, so users have little control over what is communicated between the federation partners – all personal information is stored on server side.

Open distributed repositories provide mechanisms to store specific aggregations of data (e.g. user information) in an organized way and access them for multiple purposes throughout large networks like the Internet. They are based on the concept of designing the repository as a collection of open systems that cooperate to hold a logical database of information.

Specific repositories for personal data are known as directory services. Directory services provide names, locations and other information about people and organizations. Directory information may be used for personal communication (e.g. contact information), user authentication (e.g., logins and passwords), network security (e.g., user-access rights) or various other purposes related to a specific application context.

While early network directories where mostly designed for specific applications, the first standard for an open directory service was defined in the ISO Open Systems Interconnection (OSI) model where directory functionality (directory administration, authentication and access control), is defined as ITU-T Recommendation X.500 [ITU-T, 1995]. X.500 describes a client/server architecture in which the client (directory user agent DUA) queries one or more servers (directory service agents DSA) using the directory access protocol (DAP).

Though X.500 offers a comprehensive solution for open directories, its complexity hindered widespread use – especially in the global context of the Internet. This led to the definition of a simpler TCP/IP based standard for the DAP called Lightweight DAP (LDAP) that offers comparable basic functionality and quickly evolved to the de facto directory protocol of the Internet [Yeong, W., Howes, T., et al, 1995]. While being lean, LDAP misses certain important features like full-grown security mechanisms. This fact again initiated discussions about changes and a possible return to X.500.

While directory services like X.500 and LDAP aim to be generic, domain specific (but still open) repositories try to address specific needs of dedicated application areas. An example of such a repository within the specific domain of electronic business services is Universal Description, Discovery and Integration (UDDI) which was defined by the UDDI consortium [UDDI ORG, 2002]. The UDDI repository

offers a tailored schema, protocol and architecture that allow describing and classifying business services in a natural way and accessing this information from a globally unique entry point.

Similar to the UDDI concept of a specialized global service repository, our concept could be seen as a specialized global repository for personal Identities. Among other aspects discussed in the previous paragraphs, such a tailored repository needs to combine global single sign-on mechanisms with the functionality of an open directory service. To the best of our knowledge there is currently no working system which offers authentication and directory services bundled. However, efforts in this field are done in various points. Most notably the XNS Project (eXtensible Name Service) works on protocols and XML schemas to communicate identity information over the network [XNSORG, 2002]. However, they focus on web services and therefore gain a different scope. The commercial effort PingID.com [PingID.com, 2002] tries to establish an infrastructure – again focused on web services – to share user information, but the effort is still young and unfinished, and it lacks scientific background.

5. CONCLUSION

In this paper, we showed how a general Digital Identity Infrastructure can remarkably enhance "living", i.e. communicating and co-operating on the (inter-) net. This doesn't only result from providing specific capabilities like 'single sign-on' or 'automatic form-filling' but rather from real recognition of communication partners as personal characters that can be shown to or hidden from other co-operation partners – together with all their possible aspects, fine-tune as needed in many possible fine degrees. As opposed to proprietary, application bound systems which yield only enhancements to their special area, our open Internet-wide approach to Digital Identities aims at supporting all kinds of arbitrary virtual communities with a rather generic mechanism.

Building the corresponding system infrastructure bears several technical challenges (as outlined); some of them have already been investigated in the context of the OnefC-project, others still present material for various future work. Our focus of such work will include technical aspects of identity management, identity building, identification and information exchange protocols, anonymity issues, image tracing and integration of proprietary systems.

REFERENCES

Berthold, O. and Köhntopp, M. 2001. *Identity Management based on P3P.* Lecture Notes in Computer Science Vol. 2009

Cox R., Grosse, E., et al, 2002. *Security in Plan 9.* In proc. of the 2002 Usenix Security Symposium, San Francisco.

Cranor, L., Langheinrich, M. et al 2002. *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification.* http://www.w3.org/TR/P3P/, 16. April 2002

Electronic Privacy Information Center, 2002. *Sign out of Passport!* http://www.epic.org/privacy/consumer/microsoft/

ITU-T, 1995. *Recommendation X.500, Information technology – Open System Interconnection - The directory: Overview of concepts, models, and services.*

Kohl, J. and Neuman, B. C., 1993. *The Kerberos Network Authentication Service (Version 5).* Internet Request for Comments RFC-1510.

Liberty Alliance, 2002. *Liberty Alliance Project.* http://www.projectliberty.org/

McGarty, C., Haslam, S. A., Hutchinson, K. J. & Turner, 1994. *The effects of salient group memberships on persuasion.* Small Group Research. 25, 267-293.

Microsoft Corporation, 2002. *Microsoft .NET Passport.* http://www.microsoft.com/netservices/passport/.

PingID.com, 2002. *The Identity Network.* http://www.pingid.com/.

Seybold, P. and Bock, G. E., 2002. *Our First Take on Liberty Alliance Version 1.0: Not Customer-Centric Enough!* Patricia Seybold Group, 85 Devonshire Street, 5th Floor, Boston, MA 02109-3504 USA.

UDDI ORG, 2002. *Universal Description Discovery and Integration (UDDI).* http://www.uddi.org/.

XNSORG, 2002. *eXtensible Name Service Public Trust Organisation.* http://www.xns.org/.

Yeong, W., Howes, T., et al, 1995. *RFC 1777: Lightweight Directory Access Protocol.*