

# Digital Identity Protection in Federations

December 21, 2005

Faculty supervising the research: Elisa Bertino  
Affiliation: Department of Computer Science  
CERIAS  
Telephone: 765-496-2399  
Electronic Mail: bertino@cerias.purdue.edu

PhD student : Abhilasha Bhargav-Spantzel  
Affiliation: Department of Computer Science  
CERIAS  
Telephone: 765-496-6766  
Electronic Mail: bhargav@cerias.purdue.edu

## 1 Motivation

Today a global information infrastructure connects remote parties worldwide through the use of large scale networks, relying on application level protocols and services, such as recent web service technology. Execution of activities in various domains, such as shopping, entertainment, business and scientific collaboration, and at various levels within those contexts, is increasingly based on the use of remote resources and services. The interaction between different remotely-located parties may be (and sometimes should be) based on little knowledge about each other. To support these rich experiences and collaborations, more convenient IT (Information Technology) infrastructures and systems are needed. We expect, for example, that personal preferences and profiles of users be readily available when shopping over the Web, without requiring the users to repeatedly enter them. In such a scenario, digital identity management (IdM) technology is fundamental in customizing user experience, protecting privacy, underpinning accountability in business transactions, and in complying with regulatory controls.

Digital identity can be defined as the digital representation of the information known about a specific individual or organization. As such it encompasses, not only login names (often referred to as *nym*s), but many additional information, referred to as *identity attributes* or *identifiers*, about users. Managing identity attributes raises a number of challenges, due to conflicting requirements. On the one hand, they need to be shared to speed up and facilitate authentication of users and access control. On the other hand, identity attributes need to be protected as they may convey sensitive information about an individual and can be targets of attacks like *identity theft*. By identity theft we mean the act of impersonating others identities by presenting stolen identifiers or proofs of identity. Usually, identity theft in the digital world occurs to obtain fraudulent credit or to perform crimes like accessing classified records without having the proper authorizations. Reports state that within the last twelve months about a million American adults became victims of digital identity fraud [8]. In fact, identity theft is one of the fastest growing crimes in the nation, with an estimated growth rate of 40 percent every year [3]. Please refer to Appendix B for additional information about identity theft. In cyberspace preventing identity theft is especially hard because digital information can be copied hence stolen unnoticed and also it is difficult to find or prosecute the internet thieves. Most common identity theft attacks are perpetrated through passwords cracking, pharming, phishing and database attacks. The intuitive solution of maintaining confidentiality through cryptographic techniques is inadequate when dealing with identity, as identifiers are very often public and have to be released to third parties and validated each time authentication is required or an access control policy needs to be enforced. Furthermore, identity of a user can be represented through different types and combinations of identifiers. Therefore when reasoning about identifiers, it is important to distinguish between strong and weak identifiers. A strong identifier uniquely identifies an individual in

a population, whereas a weak identifier can be applied to many individuals in a population. Whether an identifier is strong or weak depends upon the size of the population and the uniqueness of the identity attribute. The combination of multiple weak identifiers may lead to a unique identification [14]. Examples of strong identifiers are a users passport number or social security number. Weak identifiers are attributes like age and gender. This distinction is significant and is very useful in our research, because authentication can be achieved using either only single strong identifiers, group of weak identifiers, or combinations of them.

An emerging approach to address issues related to identity management is based on the notion of *federations* [2, 9]. The goal of federations is to provide users with protected environments to federate identities by the proper management of identity attributes. Federations provide a controlled method by which federation members can provide more integrated and complete services to a qualified group of individuals within certain sets of business transactions. By controlling the scope of access to participating sites, by enabling secure, cross-domain transmission of users personal information, federations can make more difficult the perpetration of identity frauds, as well as their frequency, and the potential impact of these frauds. Please refer to Appendix C for current federation system examples.

Federations are usually composed by two main entities: identity providers (IdPs), managing identities of individuals, and service providers (SPs), offering services to registered individuals. In a typical federated identity management system the individual registers with his/her local IdP and is assigned a username and password. Registration is usually based upon a face-to-face verification at some registrar office. Based on this information a registered individual can submit additional attributes and its corresponding attribute release policies, which are stored at the local IdP. The IdP is then contacted whenever the user interacts with any other SP in the federation when additional user information is needed. The IdP is in charge of sending the SP the submitted user attributes in accordance to the attribute release policies. However, current approaches to digital identity management in federated systems have several drawbacks [2, 9]. First, face-to-face registration is not always feasible, and it might be a bottleneck in large organizations. In most cases validity of strong and weak identifiers is not properly checked; individuals may use such identifiers regardless their actual temporal validity. Moreover, as we discussed, identity might not be captured by a single (or a small combination) of identity attributes. Users should be free to choose any combination of identity attributes for carrying on their on-line interactions. The identity attributes that SPs require may change from time to time, depending on the specific purpose of the interactions. Identity attributes may also have different privacy requirements.

The major drawback of current approaches is that, however, no specific techniques are provided to protect registered individuals from identity theft. The problem of identity theft is today a crucial problem because of its high financial and social costs. If no specific protection protocols are adopted, dishonest individuals can easily register fake attributes or impersonate other users of the federation. Protection from identity theft should thus be one of the main desiderata in federated digital identity management solutions.

## 2 Statement of the problem

The goal of the proposed research is to develop techniques and systems for the protection of digital identity information from identity theft. Important requirements that our solution should satisfy include the following:

1. Validity and consistency of identity information. A federated identity system should ensure consistency of the identity information shared in the federation. Although this information can only be validated by checking with actual information issuers, which can be outside the federation, the system should be able to detect identity theft based on the information available within the federation.
2. Privacy. Authentication methods should preserve individuals privacy, and enforce a need to know principle when requiring identifiers, so that only the identifiers actually required to access a service should be submitted to the SP.
3. Efficiency. The authentication methods should be efficient and require a limited number of message rounds between the SP and the user.
4. Robustness. The federation protocols should be robust, in the sense that even if an adversary is able to get the value of the strong identifiers, it should not be able to impersonate the victim in the federation.
5. Usability. As usability is one of the main aims of federations, the increased security features should not require individuals to be overloaded with computational expensive operations or require them to be involved in the authentication operations more complex or long than in the traditional federated identity systems. A security solution disrupting such simple mechanism would not be adopted.

6. Accountability. Accountability of the individual is to be ensured. It should be possible to relate individuals to actions or events for which they are to be held accountable. The system itself should be accountable with respect to the use of identity attributes; individuals should be able to check how their information has been used in the federation.

### 3 Research Plan

The research is articulated according to the following tasks:

- A) *Development of a fundamental approach to the protection from identify theft in federations*: This requires formalizing the notion of a federation and the properties that should be satisfied so as to achieve protection against identity theft. Protocols required to achieve this goal are needed to be clearly defined and analyzed.
- B) *Implementation of a prototype of the proposed approach* The protocol described in task A need to be eventually deployed for practical significance. As a first step we will implement a prototype as a proof of concept and get emperical data for efficiency and security analysis.
- C) *Extension of the basic approach with biometrics* One approach to the problem of reducing the threat of identity theft is the widespread adoption of systems of biometric identification. Biometric identification systems are automated methods of recognizing a person based on one or more physical characteristics, such as fingerprints, voice, or facial characteristics [5]. However there are several problems with a large scale and distributed management of biometric identitifiers. For example creation of a database of a particular biometric should itself be secure and possibly decentralized. Also, such database would be highly dependent on a particular vendor limiting the interoperable use of this information. Another interesting problem is the enrollment of a biometric identifier has to be with the right person else identity theft would be trivially possible. Lastly the biometric data are not completely accurate and may change overtime thus increasing the challenge of managing such information.
- D) *Development of a policy language specific to user authentication* User authentication is perhaps the single most important requirement for user privacy and security including protection against identity theft. We need a flexible yet robust approach for managing such authentication. Therefore our goal is to develop a policy language to express the authentication policies which would then be enforced by the identity system.
- E) *Application of the approach to Shibboleth* Shibboleth [2] is an initiative originated in an academic environment and its goal is to facilitate sharing of resources between research and academic institutions. We will extend this open source project to add features to protect the user information from identity theft. This would require adding new constructs and modules to the existing system.

### 4 Preliminary Results obtained by the PhD student A. Bhargav

The complex problem of identity theft requires a strong theoretical basis on which the solutions are built. To this extent we have defined the requirements of an identity system protected from identity theft and determined which tools and methodologies can be used to achieve such requirements. Instead of making strong assumptions on infrastructures and systems which are not currently prevalent, our solution shows a step by step approach of how an individual can first establish a digital identity followed by a secure and protected use of such identity. Giving user control of his or her digital identity information with respect to who can get this information and how it can be used is the key to the protection of this identity. We believe that our approach is promising for large scale deployment due to the progressive and flexible approach for enabling authentication, and the well defined security properties at each stage. We have developed a set of cryptographic protocols specifically designed to protect user attributes against identity theft. Authentication at the time of initial registration is a major concern in any identity system. We give a flexible approach for the initial registration with the help of which the user can establish an initial identity. This identity information is augmented and updated as the user interact with the identity system, thus providing a dynamic trust establishment methodology. Regarding the usage of the identity information, the main idea is to associate the different kinds of strong identifier of a user with each other and with the user's single sign-on identifier. In such way, any sensitive information of the user is not acceptable without one or more of the other associated identifying information. This effectively achieves a multi-factor authentication, which is a widely accepted desired property.

When authentication is based on multiple factors, then it is called strong authentication. We show how we can preserve user privacy without jeopardizing security of the user with the help of cryptographic techniques like zero knowledge proofs [11, 1], distributed hash tables [12] and watermarking techniques [15, 13]. An important feature of the solution we have devised is that the solution does not depend on an established public key infrastructure mechanism for the users, which is a known limitation in the practical implementation of trust management models. To the best of our knowledge this is the first time a complete solution based on composite cryptographic protocols and security tools has been proposed to solve the problem of identity theft. Effectively we have achieved task (A) from a theoretical point of view. We have submitted our most recent results to the 2006 IEEE Symposium on Security and Privacy. A comparison chart of our solution as compared to existing IdM systems is given in Appendix A.

## 5 Potential Impact and Funding Opportunities

The research being undertaken as part of this project is of high interest to various groups at Purdue and to industry. In order to maximize the potential impact, the following actions have been undertaken:

- A) A collaboration has started with the group lead by Prof. Stephen Elliot (College of Technology at Purdue) in the area of biometrics. The goal of the collaboration is to integrate biometrics techniques with our digital identity management techniques.
- B) A collaboration has started with the group from ITaP involved in the security for the TeraGrid project. The main goal of this collaboration is to determine relevant requirements concerning federated digital identity management in the area of grid computing systems. Such requirements will inform our research. The collaboration will also allow us to prepare joint proposals with ITaP dealing with security for grid computing systems and digital identity management solutions specific to federations consisting of academic institutions.
- C) The PhD student Bhargav has participated in an Identity Theft Workshop (held on July 20, 2005 in Chicago), organized by the Liberty Alliance Initiative. The Liberty Alliance includes over 150 organizations from across the globe ranging from educational institutions and government organizations, to service providers and financial institutions, to technology firms and wireless providers. As result of her attendance to the meeting, Liberty Alliance has invited CERIAS (and Purdue) to join the initiative as affiliate member. Our preliminary solutions have been discussed with various participants to the initiative and have generated a high interest from several companies participating to the initiative. We plan to actively seek funding from these companies.
- D) A collaboration has started with the database security group at the University of Texas at Dallas (led by Prof. Thuraisingham) focusing on digital identity management. As a result, a proposal is being prepared, dealing with the problem of identity theft, for submission to the NSF CyberTrust Program in February 2006.

## 6 Appendix

### A Brief Comparison of Our Results with Existing Identity Systems.

In the paper we submitted to 2006 IEEE Symposium on Security and Privacy, we proposed a novel approach for dealing with identity management in federated environment. Our main goal was to address the problem of identity theft in federations, by developing a robust and practical solution to identity management. Our approach is based on a composite protocol based on different techniques. Each of these techniques addresses a specific issue that arises when dealing with identity theft. More in details, our zero knowledge protocol makes it possible a privacy preserving proof of identity mechanism. We use distributed hash table helps in achieving consistency and robustness in the federation. Finally, watermarking techniques are required to increase usability and efficiency. Additionally, the solution we propose has some relevant features that are not present in any previous work. As shown by Table 1, our solution is the only approach that meets the desiderata required to build identity theft-protected federated identity solutions.

### B Identity Theft Numbers and Fundamental Reasons for Theft Occurance

Table 2 gives numbers which highlight the serious and rapidly growing problem of identity theft. It is reported to be the fastest growing white collar crime in US [3]. Our solution addresses the two fundamental reasons why identity theft

Criteria	Current IdM	Non Federated systems	Proposed IdM
Consistency	N	N	Y
Privacy	P	N	Y
Efficiency	Y	N	P
Robustness	N	N	Y
Usability	Y	P	Y
Accountability	P	Y	Y
Multi-factor authentication	P	N	Y

Table 1: Identity Management Systems comparison (Key: Y-Yes, N- No, P-Partial support)

occurs. Firstly, it is easy for identity thieves to assert they are someone else with the right data. Secondly, it is difficult for Identity Theft victims to prove that they are themselves once compromised. Our solution directly addresses these two concerns by building a secure proof of identity in addition to the user attribute itself.

10 million individuals affected.
\$48,000,000,000 in economic losses.
13 months before discovery.
\$5000 to \$90000 in fraudulent financial transactions.
\$600 to \$1500 out-of-pocket cost to fix.
200 to 500 hours of time to fix.
768,000 hits on Google search for "identity theft" OR "ID theft.
7,000 cases of mortgage fraud in 03; 24,000 in H104.

Table 2: Data from Gartner, ID Theft Center, FTC, FBI [6] 2004 report

## C Federation Statistics - Who's Federating

There are several benefits to federations. Federated identity can deliver several compelling benefits to organizations. Federation means that local identities and their associated data stay in place, but they are linked together through higher-level mechanisms. In Table 3 survey of some existing examples of federations is presented.

## References

- [1] I. Damgard and E. Fujisaki. A statistically-hiding integer commitment scheme based on groups with hidden order. In *ASIACRYPT '02: Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security*, pages 125–142, London, UK, 2002. Springer-Verlag.
- [2] <http://shibboleth.internet2.edu>. Shibboleth, internet2.
- [3] <http://www.bbb.org/>. Better business bureau.
- [4] <http://www.csc.fi/suomi/funet/middleware/>. Haka federation finland federation.
- [5] <http://www.epic.org/privacy/biometrics/>. Electronic privacy information center.
- [6] <http://www.ftc.gov>. Federal trade commission.
- [7] <http://www.incommonfederation.org/>. Incommon federation.
- [8] <http://www.javelinstrategy.com/reports>. 2005 identity fraud survey report.
- [9] <http://www.projectliberty.org>. Liberty alliance project.
- [10] <http://www.switch.ch/aai/documents.html>. Switchaai federation.

- [11] H. Lipmaa. On Diophantine Complexity and Statistical Zero-Knowledge Arguments. In C. S. Lai, editor, *Advances on Cryptology — ASIACRYPT 2003*, volume 2894 of *Lecture Notes in Computer Science*, pages 398–415, Taipei, Taiwan, November 30–December 4 2003. Springer-Verlag.
- [12] G. S. Manku. *Dipsea: A Modular Distributed Hash Table*. PhD thesis, Stanford University, 2004.
- [13] R. Sion, M. Atallah, and S. Prabhakar. Rights protection for categorical data. *IEEE Transactions on Knowledge and Data Engineering*, 17(7):912–926, 2005.
- [14] D. Woodruff and J. Staddon. Private inference control. In *CCS '04: Proceedings of the 11th ACM conference on Computer and communications security*, pages 188–197, New York, NY, USA, 2004. ACM Press.
- [15] X. Zhou, H. Pang, K.-L. Tan, and D. Mangla. Wmxml: a system for watermarking xml data. In *VLDB '05: Proceedings of the 31st International Conference on Very large data bases*, pages 1318–1321, 2005.

Working Group	Description
SWITCHaai Federation [10]	The SWITCHaai Federation is a group of organisations (universities, hospitals, libraries, etc.) that have agreed to cooperate regarding inter-organisational authentication and authorisation and, for this purpose, operate a <b>Shibboleth</b> -based authentication and authorisation infrastructure (AAD).
InCommon [7]	By using <b>Shibboleth</b> authentication and authorisation technology, InCommon intends to make sharing of protected resources easier, enabling collaboration between InCommon participants which protects privacy. Access decisions to protected resources are based on user attributes contributed by the user's home institution. InCommon became operational on 5 April 2005.
HAKA Federation Finland [4]	The HAKA Federation in Finland entered its production phase in late 2004. The Federation was set up in 2003, currently including 2 (of 20) universities and 1 (of 29) polytechnics as Identity Providers, and 4 service providers, including the National Library Portal (Nelli). In Finland[16], the libraries in higher education traditionally co-operate widely in licensing electronic journals. It is based on <b>Shibboleth</b> .
Microsoft, IBM, and the WS- Roadmap	In April 2002, Microsoft and IBM published a joint whitepaper outlining a roadmap for developing a set of Web service security specifications. Their first jointly-developed specification, WS-Security, offers a mechanism for attaching security tokens to messages, including tokens related to identity.
Liberty Alliance	The Liberty Alliance is a consortium of approximately 170 companies that develops specifications for federated identity management. It works on creating a single comprehensive federated identity specification. In March 2003, it released a new blueprint that described three separate specifications that can be used together or independently: First is the Identity Federation Framework (ID-FF) allows single sign-on and account linking between partners with established trust relationships. Second is Identity Web Services Framework (ID-WSF), allows groups of trusted partners to link to other groups, and gives users control over how their information is shared. Finally Identity Services Interface Specifications (ID-SIS) will build a set of interoperable services on top of the ID-WSF.
OASIS and SAML	The Security Assertions Mark-up Language (SAML) is an XML-based specification developed by the Organization for the Advancement of Structured Information Standards (OASIS). SAML provides a common language for three kinds of assertions: * Authentication assertions: declarations about a user's identity * Attribute assertions containing particular details about a user * Authorization decision assertions, which specify what the user is allowed to do at a particular site. SAML is the upcoming standard for Federated Identity Management.

Table 3: Federation Examples