Dewan, P and Dasgupta, P, "Trustworthy Identification in a Privacy Driven Virtual World, Security Topic Area (Distributed Systems Online), January 04, pg 1 - 8

Trustworthy Identification in a Privacy Driven Virtual World

Prashant Dewan
Department of Computer Science
Arizona State University
Tempe, AZ, 85281
dewan@asu.edu

Partha Dasgupta
Department of Computer Science
Arizona State University
Tempe, AZ, 85281
partha@asu.edu

ABSTRACT

The identity management of any entity in general and users in particular on the Internet is a formidable problem. This article presents an anonymity-privacy model that draws a clear dividing line between private information and identifying information of a given entity. The other contribution of this article is the abstraction of myriad forms of identity management systems (into three models) used on the Internet and in the brick and mortar world. This article also presents the challenges in identity management in peer-to-peer networks and reputation based systems. Finally, this article outlines the issues concerning secret credentials and the work being done on identity establishment with secret credentials.

Keywords

Anonymity, Privacy, Identity, Peer-to-peer, Security

1. Identity, Privacy and Anonymity

Identity can be loosely defined as a set of attributes and credentials that can be used by the members of a group, to which the entity belongs, to recognize the entity. In a similar fashion, identity can be construed as the role played by a real world entity, identification documents possessed by a real world entity, the nickname of a real world entity, or just the characteristics of a real world entity by which it is known within a group of people. For the purpose of this article, we define identity as certain information that can represent a real world entity on the Internet. The identity may or may not be traceable to the real world entity that owns the identity.

The first question we need to answer is why do real world entities need to be identified on the Internet? Isn't the revelation of an identity in a given context a breach of privacy for the identity holder? Do real world entities have a right to anonymity? Where is the dividing line between the revelation of an identity in a certain context and one's right to privacy?

A good example of such a dilemma can be scooped from the real world. In a number of countries identity systems have been challenged on the grounds of privacy supported in the constitution. The Hungarian Constitutional Court ruled in 1991 that use of the same personal identification number for multiple purposes, violated the constitutional right to privacy [1]. In 1998, the Supreme Court of Philippines ruled that the National ID system violated the constitutional right to privacy. The United States, Canada, New Zealand, Australia and the United Kingdom are some of the other countries that assign a unique id to their citizens. These ids are linked to the transactions that the citizens of the country perform, and are predominantly used for law enforcement purposes [1].

Privacy laws date back to 1361, with some references to privacy issues in the Bible[1] . The United Nations accepts privacy as a fundamental Human Right that is also explicitly recognized in the constitution of most countries. However, the United States does not recognize the right to privacy in its basic constitution. On the Internet we are concerned with information privacy and communication privacy. Information privacy includes the attributes or the history of the users, and is enforced by rules governing the collection and distribution of personal data, such as medical records. Communication privacy entails the secrecy of any data being exchanged between two users on the Internet, like E-Mail.

The entities need to be identified on the Internet in order to establish some level of trust between them. In other words, if two entities need to establish a certain degree of trust between them, they should first be able to uniquely identify each other. As we have already mentioned, the identity used on the Internet may or may not be traceable to the real life entity. Hence the trust on the identity is based on the assumption that if an identity has been used to perform good deeds, this trend will continue.

Anonymity is another highly coveted property of online transactions. Examples of online transactions are posting content on the web, chatting, sending email, or even browsing. The Oxford English dictionary defines anonymity as "the quality or state of being unknown or unacknowledged.'' If an entity is unknown to the other entities in the transaction, the entity is anonymous. On the Internet anonymity implies the execution of a transaction without revealing the real life identity of the entity. If an entity uses a virtual identity (on the Internet) to perform multiple transactions then all its transactions, can be traced back to the virtual identity, but the virtual identity cannot be traced back to the real identity of the entity. This form of anonymity is known as pseudononymity.

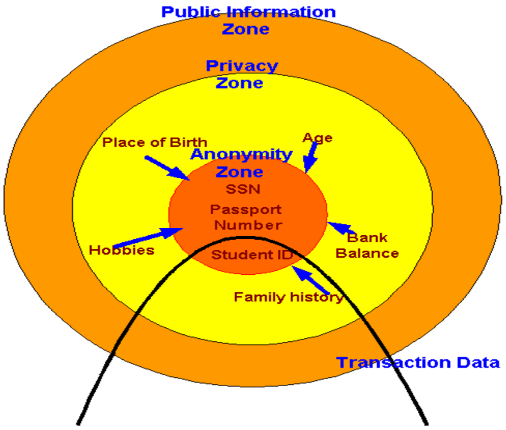

Figure 1. Privacy and Anonymity on the Internet

As shown in Figure 1, privacy and anonymity are not the same properties, although they are interrelated. Anonymity refers to the hiding the identity of an entity while privacy refers to the hiding the other attributes of an entity, which it does not want to reveal. It can be argued that identity is also just another attribute of an entity. Such an argument is correct if we consider identity as a special attribute, because many attributes in the privacy zone can be deduced from one attribute in the anonymity zone. For example, if the identity of an entity is known, its country of origin, educational background, and bank balance can be obtained. On the other hand, one needs multiple attributes in the privacy zone to deduce the identity of an entity. In other words, a subset of attributes in the privacy zone can lead to the identity of the subject. Here, it is important to note that not all subsets of the attributes in the privacy zone will provide the identity of the entity.

2. Identity Models

Identity is an essential component of trust; hence it is difficult to induce trust and anonymity in the same model. However, inducing privacy and trust in the same model of identity management is not that difficult. Anonymity-based models facilitate freedom of expression of views, but they are devoid of any accountability of the subjects. Here we assume that the entity holding the identity is legally responsible for any act performed with the identity. In the absence of ethical and moral rules binding the subjects in the model, such a model can be misused and hence deemed useless. One identity model will not fit into all scenarios. Different models have to be used in different systems based on what kinds of rights subjects exercise in a given system and the value system of the subjects. Let’s review three different models in which identities can be assigned to subjects. All the three models have their pros and cons and hence none of them can be used as a panacea for identity-based systems.

The models describe here are for citizens. These models do not apply to the governments because they have both the rights and the resources to obliterate any security built in the model. The government can spoof identities, get access to secret information and may even modify the rights associated with identities.

2.1 Model 1

In model 1, an identity is assigned to an entity by a central agency, like the government, that can be use it to perform transaction with any other subject across systems. In other words, one identity is assigned for each birth certificate produced.

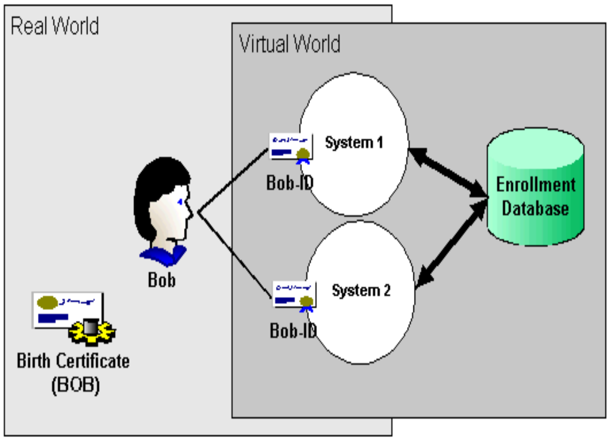

Figure 2. Model 1: Same Identity Across Systems

Subjects are authenticated by either validating their information against a central repository of authentication information or by verifying the digital signature of the authority that granted the identity certificate. The subject uses one identity to perform all his transactions on the Internet. An example of this model is the In Person Proofing at the Post Office Program, introduced by the US postal service. This system enables US residents to get a digital identity from the government. This identity can be used for any system that needs government issued identities to identify its subjects.

Another model similar to model 1 the single sign on authentication procedure and is confined to a few systems. In single sign on, an entity once authenticated, is given the rights to perform transactions on the partner systems of the authenticating system, which authenticated the subject. The subject does not need to resubmit its credentials to any of the partner systems. An example of a single sign on system is the MSN Passport. In this approach, the central agency responsible for authentication has to be trusted by the systems and the subjects. It can provide privacy depending on the amount of data collected from an entity at the time of enrollment. This service can start with an authentication mechanism for one system and can span the whole Internet by making partners. Other examples of this model are federated identity systems being advocated by Liberty Alliance[10, 11] and PingID [9].

Model 1 does not offer anonymity to the subjects. In other words, the real life identity of the subject can be easily deduced from its virtual identity because the virtual identity is mapped on to its birth certificate in the enrollment database. This model can be used to establish trust between subjects, as others can deduce the real life identity of the subjects. An important point to note here is that a traceable identity does not generate trust automatically, but is a necessary attribute for any subject for being trustworthy. It can be argued that the enrollment agency can hide the real life information of an identity, and hence provide anonymity to the subject, who is the owner of that identity. Such a system depends highly on the trustworthiness of the enrollment agency. In addition, in some countries a trustworthy enrollment agency might have to breach the trust of the enrolled subjects in order to comply with the laws of the country.

Another offshoot of model 1 is a model in which the enrollment agency does not store the mapping of the birth certificate to the virtual identity but only the list of birth certificates that have been granted a virtual identity. This model provides pseudonymity to the subject, and can establish trust by using the reputation of the virtual identity. On the flip side, this model can be misused by the subjects, as there is no way for law enforcement agencies to track down the culprit in case of a crime.

For example, an identity issued by MSN Passport may not be traceable to a real life identity because the MSN Passport does not keep a record of the real life identity, of the subject who is being issued a virtual identity. It requests for the real life information of the entity, but does not validate it. Although the IP address of the subject using the MSN Passport can be traced, it is difficult to fill the gap between the computer and the human. In other words, it is very difficult to trace the entity using a computer, from its IP address.

2.2 Model 2

In the second model of an identity system every real life entity can have multiple virtual identities. There is no way of tracing the real identity of an entity from its virtual identities. In addition, virtual identities are independent of each other, i.e., neither the action performed under any of the virtual identities nor the real identity does not impact the other identities of the subject. An example of such a system is the Yahoo chat room where one can login with multiple identities (with multiple clients), and all the actions performed by each virtual identity are totally independent of the other virtual identity or the real life identity. There is no method by which any of yahoo identities can be traced back to the real life subject. A session of a virtual identity can be traced back to the IP address, from where the entity is logged on the chat room. As already mentioned, the gap between the machine and the human is too big to fill. Hence the probability of tracing the real entity down is nil, given the fact that the entity does not disclose any other information than what is required by the chat room. Yahoo requests for the real life information of the entity, but does not validate it.

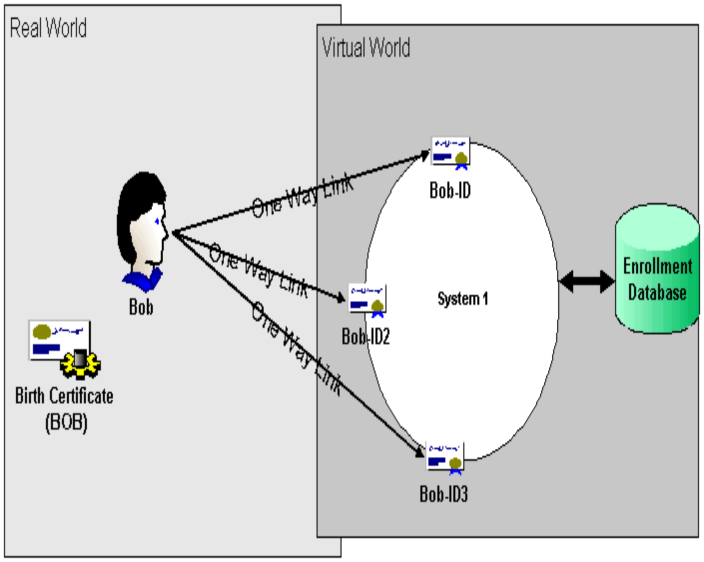

Figure 3. Model 2: Multiple Untraceable Identities in One System

This model provides a high degree of anonymity and privacy and is effective in chat rooms where the rights of the identities are confined to expression of views. The rights of the entities in these chat rooms do not give them any leverage to break any laws. Although entities can break the law by disclosing confidential information without being identified, they do not enjoy any credibility due to their anonymity. Hence not many peers will have confidence on the information gathered from a chat room. The entities can express views while anything inappropriate is censored by the system. Hence yahoo chat rooms are examples of a low trust model, which work well.

Currently, the information extracted from chat rooms is not considered credible. Besides, there is no motivation for the members of a chat room to abstain from disseminating false information in the chat rooms. A reputation-based system, where the reputation of the members in a chat room is based on the accuracy of the information that they provide injects trust in this system. Reputation can also motivate the members not to cheat or disseminate incorrect information. A reputation-based system is examined in Section 3 of the article.

The proposed model will not work in applications where the members have high stakes and high rights due to the inherent lack of trust in the system. Once an entity gets an identity in the system, it can use its new power obtained by the virtue of its identity in the system, to do certain unlawful things, such as disclosure of certain confidential information of the company, or cheating in certain stock transactions. This model can be used in high stake transactions if trust is injected between peers.

2.3 Model 3

In the third model, an entity has multiple virtual identities, which are linked to the real life identity (core identity) of the entity. Any action performed with the virtual identities can be traced back to the real identity. An example of this model is the Social Security model of the United States. The US government provides one SSN for each birth certificate- which boils down to one core identity to each real life subject. A real life subject can possess multiple secondary identities like a school id, a driver's license, a passport etc. All the secondary identities are mapped onto the core identity. This model does not provide any anonymity to the subject; it only provides selective privacy. The model however helps in enforcing laws by enabling the law enforcers to trace the real life subject, if any of the identities is used to commit any illegal activities. This model allows subjects to only divulge the necessary credentials for a given transaction. For example, for most of the transactions performed in the school, an entity can use its student id and does not have to reveal its SSN. One of the inherent problems with this model is the allocation of rights and responsibilities to the identities of the peer. In other words, a peer might find it difficult to choose an identity for a given transaction from its set.

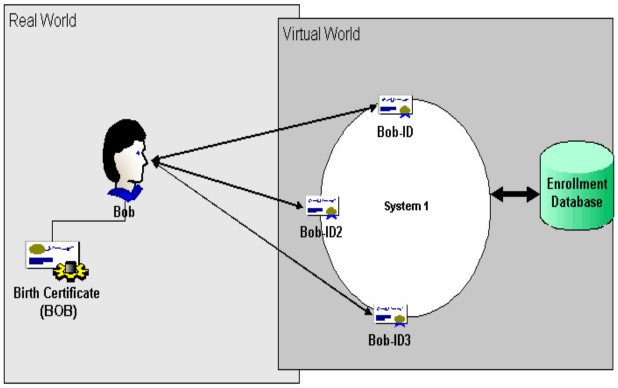

Figure 4. Model 3: Multiple Traceable Identities in One System

It might be argued that laws are used to preserve user privacy. Such laws are enforced by the government, which binds data collection agencies to protect data from stealth. The flip side of these laws is that most privacy laws are overruled when the judiciary issues a subpoena. In other words, a user's private information is no longer private if requested by the judiciary. Hence this model provides selective privacy, a high level of trust, but cannot guarantee anonymity. The model can be used in high stake applications, like stock markets, but it suppresses the right to express unpopular views. An entity identified by its social security number will make sure that it does not break the rules, so that it can be trustworthy, but it will never express its true views if its views are likely to raise furor among the community.

To summarize the above models, model 3 has the potential to provide the highest level of inter-peer trust. As previously mentioned, the identification of the subject is a necessary component of the trust between two subjects. In addition it provides the strongest link between the virtual and the real life identities, thus providing a strong motivation for the entities to behave in a trustworthy fashion. Most countries support the privacy of the users in some form or another in their legislature. This model may not be implemented without external mechanisms to protect user privacy. Model 2 provides anonymity and privacy and can be used with reputation mechanisms for medium stake transactions. Medium stake transactions are associated with small enough risk, which the transacting parties can absorb the damages. Model 1 allows real world entities to possess only one identity across systems and does not guarantee anonymity.

3. Peer-to-Peer and Identity

The above three models make the implicit assumption that there is always a trusted third party to facilitate the transaction. Although it is a reasonable assumption in small systems like corporations and schools, this assumption cannot be made in case of large systems, like the Internet. It can be argued that Verisign is a trusted third party on the Internet; however in the recent past there have been instances where the trust put in Verisign was breached [7]. Hence even Verisign cannot be trusted unconditionally. On the Internet, it is very difficult to find a third party which is trusted by two strangers. The second problem with this assumption is that if the trusted third party is compromised, the models fall apart. For example, if the enrollment agency in model 1 or model 2 gives the same identifier to multiple people or if the trusted agency reveals the SSN in model 3 for some money, these models will fall apart.

Self-certification by peers [5] can be used by entities to generate their own identities and obviate any dependency on the third party. The other peers in the network can use their past experience with the entity to authenticate and, subsequently, transact with the entity. The past behavior of the entity is usually attached to its identity in the form of reputation. Some of the decentralized reputation systems developed by researchers are RCert[8], P2PRep[3] and EigenTrust [6] . Reputation systems are based on the premise that an entity will not share its identity with others, i.e., will not allow others to use its identity for the fear of abuse of its trustworthiness. An excellent example of a server assisted reputation system is e-Bay. In e-Bay a central server stores all reputation data. All the subjects registered in the e-Bay database trust the central server.

The use of self-certification with reputation raises some important questions. The first question being, if an entity is capable of generating its own identity, then how do we stop an entity from shunning an identity with a bad reputation and generating a new identity that has no reputation? It can be argued that once a good reputation is associated with the identity, the entity would be motivated to hold on to that identity and hence abstain from participating in ‘illegal’ activities. This is only partially true, because an entity can generate as many identities it wants, improve their reputations and perform one high stake, bad transaction and shun the identity. Another possible solution to this problem is the creation of a new identity, which can be made difficult and hence provide motivation to entities to abstain from discarding old identities. Some of the techniques that can be used to make the generation of new identities difficult include micropayments [4] and mandatory references.

Micropayments force the entities to perform a certain amount of work before generating an identity. For an entity with a small amount of resources, it might be difficult to give micropayments, but entities with huge resources can easily surpass this hurdle. Another technique for enforcing constraints on entities is the use of references. The entities have to find at least N references with M reputation to generate a new identity in the system. This technique can stop the entities from generating new identities whenever they desire. A mechanism is also needed to stop identities with good reputation from giving references without any consideration. The entities can be discouraged against providing random recommendations by making them pay a cost for every reference they provide in the form of a (slight) reduction in reputation or monetary costs or micropayments for every reference they give.

The next problem is the determination of the value of N and M for a given system. Unfortunately, there are no easy answers to this problem, because a higher level of N and M increases the security of the system, by forcing entities to stick to their old identities and hence maintain good associated reputation. On the other hand, the reference restriction reduces the usability of the system because it might not be possible for all entities to find out N identities in the system without joining the system. If the system values of N and M are reduced, the system becomes more accessible but the security of the system is compromised. Figure 5 shows the tradeoff graphically.

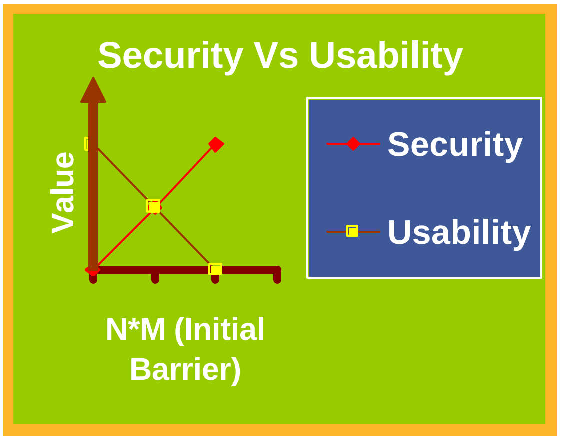

Figure 5. Security Vs. Usability

4. Identity with “Secret Credentials”

Another assumption the above three models make is that the subjects would readily disclose their credentials. In other words, it is assumed that each entity possesses public credentials, and the revelation of these credentials does not breach privacy. This assumption might not be true in certain scenarios. For example, a national passport might be a sufficient credential for a person (say Bob) to prove his citizenship, but he might not like to show his passport to another person (say Charles), unless the second person shows proof of identitfication.

The other view of the same problem is that, unless the other person (Bob) proves his trustworthiness by showing his credentials, the first person (Charles) might not disclose his credentials. If the credentials of the other party are public, like employee badges, then this problem becomes trivial. The entity with the credentials in the public domain discloses its credentials before the other party.

In scenarios where both the parties have underlying policies requiring the other party to prove possession of a minimum credential before disclosing its own credential, this might lead to a chicken and egg situation. For it has been shown, that in any bilateral transaction, the party who acts first has a higher chance of losing in the transaction [4]. Therefore at least one of the parties should possess some public credentials, i.e., a credential for which it has not minimum requirements from the other party in the transaction. The sequence of disclosure of such credential has been extensively researched by the research group of Winslett, under the heading, Automated Trust Management [12]. Automated Trust Management techniques can be used to arbitrate the disclosure of credentials when none of the interacting entities have public credentials.

Automated Trust Management assumes that the users have a set of credentials and all services have policies, which are satisfied by a specific set of credentials. The credentials also have policies that they can disclose only if the other party meets a certain criterion. Hence both parties start disclosing credentials starting with unprotected credentials (e.g. name, age, location), i.e. credentials that do not have criteria for disclosure. This protocol continues till the user discloses the required credentials for the service. From thereon, the user can use the service.

IBM’s project Idemix[2] satisfies some of the needs of anonymity and discretionary disclosure of credentials. It is based on the principle that only the information that is absolutely necessary for the execution of a transaction should be disclosed to the other party. For example, if Alice needs a rental car and has a driver’s license, then the rental company only needs to verify that Alice possesses a driver’s license. The rental company does not need to know the name and address of Alice. This is true except in case that Alice gets involved in an accident with a rental car, or if she tries to run away with the car. In that situation, the rental company will need additional information about Alice. Hence a safety procedure is implemented by which the rental company acquires the name and address of Alice in certain untoward scenarios. Here, it is interesting to note that in order to impose constraints on subjects enjoying anonymous multiple identities, a back up procedure is needed in certain situations to trace those identities to the real life subjects.

The National Electronic Commerce Coordinating Council (NECCC) is working on the developing an infrastructure for the E-government. NECCC, in its white paper, has proposed various identity management approaches, such as a single identity per person, each person generating his own identity. The identities are generated by the states or non governmental organizations. Interestingly the state of Massachusetts grants two identities to each real life entity. One identity is meant for the US government while the other identity is meant for the Commonwealth of Massachusetts. Both identities can be traced back to the same real life subject therefore the two identities issued to one person are actually one identity per system. Hence they still follow one identity to one real life subject. The Iowa state government is working on an Identity-Security project to link multiple identification documents to one identity. This identity is granted on the policy of one identity for one birth certificate.

4.1 Conclusion

Human identification on the Internet is a difficult problem that does not have any silver bullet. Different models have to be chosen for different systems, by estimating the rights to be imparted to the identity holders, their duties and the amount of explicit order that needs to be enforced. In addition, the same definition of privacy cannot be used across systems without considering the needs of the entities participating in the system.

References

[1] David Banisar and Simon Davies. *An International Survey of Privacy Laws and Practice* [Electronic]. Global Internet Liberty Campaign, [cited July 09 2003]. Available from: http://www.gilc.org/privacy/survey/

[2] Jan Camenisch and Els Van Herreweghen.*Design and Implementation of the Idemix Anonymous Credential System*. IBM Research Division.2002

[3] Ernesto Damiani, De Capitani di Vimercati DEA, Stefano Paraboschi DEI, Pierangela Samarati DTI and Fabio Violante DEI. *A Reputation-Based Approach for Choosing Reliable Resources in Peer-to-Peer Networks*. Paper presented at the Conference on Computer and Communications Security (CCS 02), Washington, DC, USA 2002.

[4] Chrysanthos Dellarocas, Sloan School of Management. and Center for eBusiness@MIT. *Building Trust on-Line : The Design of Reliable Reputation Reporting* ,Ebusiness@MIT Working Paper ; 101. Cambridge, Mass.: MIT Sloan School of Management, 2001.

[5] Prashant Dewan, Austin Godber and Partha Dasgupta*. A Self-Certification Scheme for Managing Reputations in Peer-to-Peer Networks*. Arizona State University.Technical Report.2003

[6] Sepandar D. Kamvar, Mario T. Schlosser and Hector Garcia-Molina. *The Eigentrust Algorithm for Reputation Management in P2P Networks*. Paper presented at the Twelfth International World Wide Web Conference 2003.

[7] Microsoft. *Microsoft Security Bulletin Ms01-017: Erroneous Verisign-Issued Digital Certificates Pose Spoofing Hazard* (V2.3) [Newsletter]. Microsoft Corporation, June 23,2003 [cited July 23,2003]. Available from:http://www.microsoft.com/technet/ treeview/default.asp?url=/technet/security/bulletin/MS01-017.asp

[8] Beng Chin Ooi, Chu Yee Liau and Kian-Lee Tan. *Managing Trust in Peer-to-Peer Systems Using Reputation-Based Techniques*.
Paper presented at the The 4th International Conference on Web Age Information Management, Chengdu 2003.

[9] PingId. *Pingid* [Electronic]. PingId.com, October 15, 2003 [cited December 04,2003]. Available from:http://www.pingid.com/

[10] Liberty Alliance Project. *Identity Systems and Liberty Specification Version 1.1 Interoperability*. Liberty Alliance Project.2003

[11] Liberty Alliance Project. *Introduction to the Liberty Alliance Identity Architecture*. Liberty Alliance Project.2003

[12] Marianne Winslett. *An Introduction to Automated Trust Establishment*. Paper presented at the Workshop on Credential-Based Access Control, Dortmund 2002.

Acknowledgements

I want to acknowledge Austin Godber whose comments have been invaluable for my comprehension of identity systems in particular and security in general. I also want to acknowledge all the reviewers of these articles whose comments have made this article more informative and useful.

Biography

Prashant Dewan is a doctoral candidate in the Department of Computer Science in the Arizona State University, Tempe, USA. He completed M.S. in Computer Science in May 2002 from the same university. He is using reputation techniques to inject trust into peer-to-peer and ad hoc networks, as a part of his work towards his doctoral dissertation. His research interests are operating systems, distributed systems and security. He is a student member of ACM and IEEE. His research can be found at www.public.asu.edu/~dewan

Partha Dasgupta received his Ph.D. in Computer Science in 1984 from the State University of New York. His prior education was at the Indian Institute of Technology. He is on the faculty of Arizona State University, and teaches courses in Applied Cryptography and Distributed Operating Systems. He has over 15 years of research experience in distributed Computing, Operating Systems, Security and Networking. His research is funded by NSF, DARPA, AFOSR, Microsoft and Intel. His research can be found at cactus.eas.asu.edu