Association for
Computing Machinery

Advancing Computing as a Science & Profession

Testimony before the House Committee on Ways and Means Subcommittee on Social Security on

Protecting the Privacy of the Social Security Number from Identity Theft

21 June 2007

Statement of
Ana I. Antón, Ph.D.

Associate Professor
North Carolina State University

Director
ThePrivacyPlace.Org

On Behalf of USACM (the US Public Policy Committee of the Association for Computing Machinery)

Introduction

Thank you Chairman McNulty and Ranking Member Johnson for the opportunity to testify.

I am an associate professor at North Carolina State University in the Department of Computer Science in the College of Engineering. I am the director of a privacy research center named ThePrivacyPlace.Org, with collaborating faculty and students at NC State University, Purdue University and Georgia Tech. In addition to my role as a faculty member, I serve on several industry and government boards of technical advisors, including the Department of Homeland Security's Data Privacy and Integrity Advisory Committee. A brief biography is in Appendix A.

This statement represents my own personal position as well as that of the Association for Computing Machinery's (ACM) Committee on U.S. Public Policy (USACM), of which I am a member of its Executive Committee. ACM is a non-profit educational and scientific computing society of more than 84,000 computer scientists, educators, senior managers, and other computer professionals in government, industry, and academia, committed to the open interchange of information concerning computing and related disciplines. The Committee on U.S. Public Policy acts as the focal point for ACM's interaction with the U.S. Congress and government organizations. It seeks to educate and assist policy-makers on legislative and regulatory matters of concern to the computing community. (See http://www.acm.org and http://www.acm.org/usacm.)

My statement today highlights two key policy and technology issues:

First, the Social Security number (SSN) should not be used as an identifier or authenticator. Such conflicting uses of SSNs confuse the role of identity and authentication, making the SSN valuable to for stealing someone's identity or committing fraud. Furthermore, because SSNs are so readily available, they are not an adequate means of identification or authentication.

Second, steps must be taken to reduce the use and exposure of SSNs. Reducing the use of SSNs helps protect individual information. The display of SSNs on ID cards and in public records should be prohibited, and SSNs should be redacted from existing public records. From a technical standpoint, we should require government and private entities to secure or encrypt records or documents containing SSNs in storage and during transmission.

Overview

Identity theft is no longer a rare crime impacting few and escaping the general public's attention. Since 2003, more than 36 million Americans have had their identities stolen[1] and since February 2005 over 155 million personal records[2] have been compromised, causing serious harm to those affected by these thefts and exposures. Massive data breaches, such as last year's stolen laptop containing the Social Security numbers (SSN) of 28 million veterans, are increasingly commonplace and will soon cease to be front-page news. With increasing public awareness of this epidemic, Congressional attention has followed, with no less than eight different committees conducting oversight hearings and proposing numerous bills.

Identity theft is a form of fraud that depends on two different factors, 1) information being improperly acquired, and 2) that information being used to facilitate fraud. The improperly acquired information is the enabler, but it is the fraud that does the damage.

A number of factors enable identity theft, but two key ones stand out for this committee's consideration. First, the use of the SSN is so widespread that it is a de facto national identification number. Businesses and government agencies collect the SSN to identify and then authenticate individuals. This has made it the primary instrument for stealing an individual's identity or creating a "synthetic identity," which is a new identity cobbled together using personal information from several sources. It is the key that unlocks access to credit, banking accounts and various other services for criminals.

Second, current computing technologies enable the collection, exchange, analysis, and use of personal information on a scale unprecedented in the history of civilization. As computing technology and storage continue to advance and become cheaper, and as more uses are found for information about people, we can only expect these collection activities to increase. Whereas paper records with personal information, including SSNs, used to require some effort to find, copy and disseminate, the spread of inexpensive computing technology has made it much easier to find, use and exploit such information – for good and for bad. Moreover, SSNs provide a way to easily link data records that should not be easily linkable — a key point in enabling data theft. Reducing the use of SSNs will make it more difficult for thieves to, for example, tie together medical and investment records.

Collectively, these trends suggest that it is a critical time for Congress to act to strengthen the privacy of Social Security numbers. The use of SSNs is widespread, and no amount

---

[1] Javelin Strategy and Research, "2006 Identity Fraud Survey Report", January 2006, http://bbb.org/Alerts/article.asp?ID=651 and Javelin Strategy and Research "2007 Identity Fraud Survey Report, data obtained from Privacy Rights Clearinghouse, "How Many Identity Theft Victims Are There? What Is the Impact on Victims?", accessed June 12, 2007, http://www.privacyrights.org/ar/idtheftsurveys.htm

[2] Privacy Rights Clearinghouse, "A Chronology of Data Breaches", accessed June 12, 2007, http://privacyrights.org/ar/ChronDataBreaches.htm.

of technology or law can "put this genie back in the bottle." We can, however, construct sensible policies, combined with new business procedures, and deploy technology using best practices for privacy protection. Such measures will help protect SSNs and move us away from relying on them as the key to unlocking one's identity.

In this testimony, I will touch on several issues that describe the nature of the privacy and security problems with SSNs. I will then discuss alternative approaches to managing identity and plausible technical solutions for protecting privacy. Finally, I will discuss some recommendations for this committee to consider as it moves forward with efforts to protect the privacy of SSNs.

Use as an Identifier and Authenticator

The U.S. government issues Social Security numbers to track taxes and benefits. Originally, the number was connected only with the Social Security program. Over time, governments and other entities have expanded the uses of the SSN. Issued to individuals for nearly three quarters of a century, each SSN is intended to be a unique number that people keep for life. The problem we face today stems from the fact that the SSN is so convenient for tracking individuals across public and private records that it is often used as both an identifier and an authenticator.

An *identifier* is a name or other label that can be used to uniquely select a particular person within a specific group or context. For example, my SSN identifies me within the group of U.S. Social Security participants. But someone who knows my SSN is not necessarily me. Many other people in many contexts have valid access to my SSN.

Authentication is the process of verifying that an identifier is valid and associated with a particular identity. There are three traditional categories of authenticators: knowledge-based ("what you know," e.g., a password), object-based ("what you have," e.g., an RFID token or a driver's license), and ID-based ("what you are," e.g., a biometric such as a fingerprint).[3] There are strengths and weaknesses in each form of authenticator; these are discussed in more detail in USACM's short tutorial on authentication, attached as Appendix B.

Companies and government agencies rely on the SSN as an identifier because it gives them some added degree of precision when disambiguating individuals. For example, when distinguishing between 20 different Sally Smiths in a database, the SSN is presumed to be unique and indicates exactly one individual. This holds true whether Sally enters her name as Sally Ann Smith, S. Smith, or uses her married name from one of several marriages. Clearly, this identification role is valuable to ensure that records are not mixed and duplicate records are not created. However, privacy problems ensue if access to those same personal records are provided when someone claiming to be Sally provides the SSN as an authenticator. Her family members, former college roommate and professors, employer, banker, former spouses and others likely know her SSN. So does

[3] O'Gorman, L. Comparing Passwords, Tokens, and Biometrics for User Authentication. *Proceedings of the IEEE,* Volume 91, pp. 2021-2040, 2003.

anyone with access to Sally's records at any organization that uses her SSN to identify her out of all the other Sally Smiths.

From a technical standpoint, the problem with SSNs is that some entities are using them as an identifier, others are using them as an authenticator, and yet others are using SSNs as both an identifier and an authenticator.[4] For example, professors applying for some Federal grants via the Internet enter their social security numbers, last names, and a password to identify themselves. In this case, the SSN is used as an identifier and the last name as a form of secondary authentication. As another example, use of the SSN when opening a bank account is as an identifier, to indicate who is responsible for taxes. As a last example, if you call your credit card company to increase your credit limit they may ask for your SSN to authenticate you. These conflicting uses confuse the role of identity and authentication and thus make SSNs much more valuable for stealing someone's identity.

Remotely conducted business transactions that rely on the SSN as an identifier and authenticator are particularly risky. If someone is contacting his brokerage via the phone or Internet to purchase stock against a credit card, the brokerage needs to authenticate the identity of the customer as the holder of the credit card and the brokerage account. The customer also needs to authenticate that he is communicating with the real brokerage, and not with a criminal seeking to steal his credit card information. Furthermore, the authentication needs to be performed in a way that someone eavesdropping on the transaction cannot then masquerade as either party for any other operation. Knowledge of a SSN (or any other universal identifier) is not sufficient to reliably authenticate any party in this transaction, but this use is commonplace.

Ubiquity of SSNs

SSNs are so widely used that they are a de facto national identifier. Commercial entities and government agencies rely on their use, some states have them on driver's licenses, employers use them as medical plan IDs and employee IDs, and many colleges have long used them as student IDs. (In recent years universities have moved away from the use of SSNs). The military also uses SSNs as military serial numbers and has them exposed on the Common Access Card that every member the military carries for identification[5].

In testimony before Congress[6] in 2004, the General Accounting Office found that an estimated 42 million Medicare cards displayed entire 9-digit SSNs, as did approximately

---

[4] S.L. Garfinkel. Risks of social security numbers. *Communications of the ACM* 38(10), pp. 146, October 1995.

[5] B. Acohido & J. Swartz. "Military Personnel Prime Targets for ID Theft," *USA Today*, June 15, 2007, accessed June 18, 2007, http://www.usatoday.com/tech/news/computersecurity/infotheft/2007-06-14-military-id-thefts_N.htm?csp=34.

[6] Government Accountability Office, "Social Security Numbers: More Could Be Done to Protect SSNs", March 20, 2006, http://www.gao.gov/new.items/d06586t.pdf

8 million Department of Defense (DOD) insurance cards and 7 million Department of Veterans Affairs (VA) beneficiary cards.

As we have moved from paper-based systems to digital ones, the exposure of SSNs has increased. SSNs are often used as identifiers on public documents, such as property deeds. Before the widespread use of digital technologies, databases and networking, these documents represented little threat of being a source of large-scale identity theft. However, now such documents are digitally scanned and incorporated into massive databases often held by data brokers that share or sell this data. This facilitates identity fraud in dealing with other organizations that use knowledge of the SSN as a form of authentication. This secondary use and electronic transmission of this data creates an array of opportunities for the data to be intercepted and misused.

Data Security

Two weeks ago, before this subcommittee, Dr. Peter Neumann[7] (an expert on privacy, security and trustworthy computing issues) discussed the security and reliability vulnerabilities of the existing computing infrastructure and the poor current state of practice in building trustworthy systems. These vulnerabilities can and have led to the exposure of personal information, particularly SSNs. Nearly two-thirds of the security breaches from January 1, 2005 to June 11, 2007 resulted in the exposure of SSNs.[8] (There were more than 600 publicly reported data breaches during time period and, of these breaches, more than 400 potentially exposed SSNs.[9])

Creating complex systems that are dependably trustworthy (secure, reliable, survivable in the face of many adversities) remains a grand challenge of computer science. Further, because many of the privacy problems are related to total systems, encompassing computers, communications, people, and procedures, they cannot be adequately protected by technological approaches alone. While better computer security is a worthwhile technical and policy goal, these factors suggest that a better approach is to move away from reliance on SSNs – at the very least, to avoid depending on them as means of authenticating identity.

Alternative Identifying Techniques

Reducing the use of SSN as both an authenticator and identifier will better protect privacy and security; however, the question is, what other techniques can we use that better protect identity?

---

[7] Testimony of Peter G. Neumann For the Congress of the United States House of Representatives Committee on Ways and Means Subcommittee on Social Security Thursday, June 7, 2007, http://www.acm.org/usacm/PDF/EEVS_Testimony_Peter_Neumann_USACM.pdf

[8] Privacy Rights Clearinghouse, "A Chronology of Data Breaches", accessed June 12, 2007, http://privacyrights.org/ar/ChronDataBreaches.htm.

[9] Ibid.

A study by the American Journal of Public Health detailed the effectiveness of combinations of data used as identifiers instead of the SSN[10] It found the SSA National Death Index could be used to identify people by certain data points (First Initial, Last Name, Day of Birth, Month of Birth, and Year of Birth) as accurately with or without the SSN. This means that adding the SSN falls within the margin of error when searching with only the first initial, last name and birth year. Adding full name, full birth-date, and possibly some other information such as birthplace would likely provide as complete and unique identification as the SSN.

We should not, however, replace one enabler of ID theft with another. Many of these data points are more readily available to strangers than SSNs. Thus, it is important that they not be used to *authenticate* identity for anything of value. Clearly, context for determining identity matters, but the larger issue is that we should move away from reliance on any system that creates universal identifiers, and especially from systems that use knowledge of those identifiers as authenticators of that same identity. It is more secure to use multiple factors to confirm an individual's identity. For data aggregators looking for a universal identifier, it may be more convenient – perhaps even more profitable – to work on ways to effectively sort and target their analyses to confirm that the information they collect is appropriate for their intended uses.

Designing Systems to Protect SSNs

Databases containing personal information often employ the SSN as the "primary key" or common identifier. This exacerbates the problems I have addressed as it presents yet another vulnerability by making it easier to match records from disparate data sources. Replacing SSNs as the primary key in these systems is not a large technical hurdle in most cases. Random 9-digit numbers would work for virtually every company in the U.S. and would not be difficult to generate. We can better protect individual privacy by using different, random numbers in each company database. This would prevent someone from easily correlating the personal data an individual held in several of those databases.

The technical challenge is in replacing the SSN in all those databases. First and foremost, if an SSN is being used in a database as an index, then replacing it will require updating the index. Indexes generally exist to make common selection operations faster in databases. The worst-case scenario would be a database that has a great many transactions per day and little to no downtime to update the index. For large, always-on databases, replacing the SSN and updating the index can be a difficult process, but it is possible.

The cost of cleaning up records in commercial databases is likely far from prohibitive for all but the smallest of commercial entities. Usually smaller entities have less expertise, time and money to scrub their records of SSNs. But, the smaller entities are also the least attractive targets for identity theft.

---

[10] B.C. Williams, L.B. Demitrack, and B.E. Fries. "The accuracy of the National Death Index when personal identifiers other than Social Security number are used," *American Journal Public Health.* 82(8): 1145-1147, August 1992.

The most vulnerable databases are employee records, which are typically not as secure as other business databases both because of insider threat and because they are often overlooked in audits of business assets.[11]

Universities are increasingly discontinuing the use and storage of SSNs whenever possible, replacing them with university-specific ID numbers (i.e., a random 9-digit identifier that is not the SSN). In cases where SSNs must be collected and stored (e.g., for employment and financial aid), the SSN is no longer being maintained as the common identifier or primary key. In addition, universities are actively adopting technologies to protect SSNs. For example, devices containing SSNs are being protected by a password-based security system using encryption. This voluntary approach can serve as a model for other industries in the U.S. to similarly move away from the use of SSNs as primary identifiers.

Recommendations

There are several actions that can be taken to protect the privacy of SSNs. I present two sets of recommendations. The first set of recommendations address the purposes for which SSNs are used. The second addresses how SSNs are stored and transmitted.

Recommendations on the Purposes and Uses of SSNs

- No bank, credit agency, government agency, or other entity should verify the identity of a person based on weak authenticators such as knowledge of an SSN or mother's maiden name. Instead, they should require stronger authentication of identity when conducting business. Strong authentication can initially be provided by a passport, military ID or license with a photograph.[12] Once that is established, a secondary authenticator such as a secret, shared password or PIN can be used for subsequent transactions. While this requires additional effort, it provides extra layers of security, and should help assure the public that the security and privacy of their information is being taken seriously.

- Universities are voluntarily moving away from SSNs in an attempt to reduce their liability under the Family Educational Right to Privacy Act (FERPA) if SSNs are accidentally exposed. This has also led to improvement of their database security. Congress should consider making private institutions financially liable in a manner similar to universities: Consumers would have a civil right of action if their SSNs or other personal information is collected or exposed without a valid business need-to-know, whether intentionally or inadvertently.

---

[11] Stephanie Armour. "Employment records prove ripe source for identity theft," *USA Today*, January 23, 2003. http://www.usatoday.com/money/workplace/2003-01-23-idtheft-cover_x.htm

[12] There are other dangers to taking this approach to a logical conclusion: having some other form of national ID. We do not support this alternative.

- There should be no penalty or discrimination against someone who will not provide an SSN when conducting business, except where required by law to disclose that information. Moreover, entities that collect SSNs should provide notice that there will be not be a penalty for withholding the SSN. Within this context, it should be an unfair trade practice for private entities (e.g., utilities) to penalize or refuse to transact business with someone who declines to provide an SSN that is not required by law. Coupled with the previous recommendation, this provides a strong incentive to move away from the SSN. It is also consistent with advice from the Federal Trade Commission on protecting oneself from identity theft.[13]

Recommendations for Protecting SSNs in Storage and During Transmission

- The display of SSNs on ID cards and in public records should be prohibited. Further, SSNs should be redacted from existing public records. Redacting SSNs is, admittedly, complex because computers (and the Internet) are not good at "forgetting" things. Once something is stored in a computer, erasing it everywhere it is stored is difficult.

  The National Security Agency has posted guides[14] on how to redact data, but the reality is that redaction has a bad reputation as a workable solution. Paper redaction techniques such as covering a name or diagram with a blackened area do not work well in digital files. There are also so many different formats for digital files that the techniques that do work are generally not portable among formats. However, ChoicePoint, as one of its newly adopted business practices in the wake of its landmark $16 million settlement for endangering the privacy of over 160,000 consumers in 2005, is now redacting SSNs and other personal data from the public records that it provides to its clients. This is a welcome development and one that could be required at other companies, especially data brokers and credit bureaus, as a matter of public policy.

There are other steps that can be taken to reduce the use and exposure of SSNs. They include:

- Require transmission of records or documents to be secure or encrypted if they contain SSNs and other personally identifiable information. Even basic Secure Socket Layer (SSL) encryption would help reduce the incidence of exposure and minimize the damage. SSL technology is readily available and is commonly used by almost every reputable e-commerce site. Properly encrypted data is usually protected even if it is stolen (e.g., on a laptop disk or thumb drive).

- Require electronic security for files and devices containing SSNs. Each instance of access to SSNs in databases should be logged for audit purposes and require a

[13] Federal Trade Commission, "ID Theft: What's It All About?", June 2005. http://www.ftc.gov/bcp/conline/pubs/credit/idtheftmini.pdf

[14] See both http://www.fas.org/sgp/othergov/dod/nsa-redact.pdf and http://www.nsa.gov/snac/

secure session. These measures insert a level of security and accountability into the maintenance of these databases. With proper access controls (where individuals who access the database are controlled, recorded, and their specific access limited to a minimal number of records), individuals can be held accountable for breaches and exposures of SSNs that can be traced to a specific access of a database.

- Eliminate the SSN as primary key in databases containing SSNs. A primary key uniquely identifies each record in a database table. Instead of using the SSN, the primary key should be a unique number generated by the database management system. Universities and other large institutions have made this transition, and their example should be encouraged in other agencies, companies and organizations.

- USACM has a set of recommendations for enhancing the security, privacy and accuracy of personal data kept in databases. Those recommendations are attached as Appendix C. We encourage the committee to consider how these might be integrated and supported by any legislation crafted to protect SSNs and related personal information held either by government or by private organizations.

Conclusion

The increasing incidence of identity theft and data breaches requires that all entities, public and private, take steps to better protect information. Where this subcommittee can help is in encouraging a reduction in the use of the SSN in commercial and government transactions and in public records. The SSN is used much more than is necessary and the ubiquitous presence of SSNs makes them ideal targets for identity theft. Because these numbers are so readily available, they are not an adequate means of identification or authentication as many consider them to be. Reducing the use of SSNs helps protect individual information and encourages businesses, government and universities — some of which have already started to make changes — to move towards other means of identification. These alternate means in many cases can be as accurate, and more secure, than using the SSN. Increased privacy and security is not only a public good, but may make people more comfortable about conducting business online.

Acknowledgments

I am particularly grateful to Cameron Wilson (ACM Director of Public Policy), David Bruggeman (ACM Public Policy Analyst), Eugene H. Spafford (USACM Chairman, and Professor at Purdue University), Travis D. Breaux, Laurie A. Jones, Aaron K. Massey, Paul N.H. Otto, and many other members of ACM and USACM for their generous help in my preparing this testimony.

Appendix A – Biographical Information

Dr. Annie I. Antón is an Associate Professor of Software Engineering in the College of Engineering at the North Carolina State University. She received her Ph.D. in Computer Science in June of 1997 with a minor in Management and Public Policy from the College of Computing at the Georgia Institute of Technology in Atlanta. Her thesis co-advisors were Dr. Peter A. Freeman and Dr. Colin Potts. She received a BS in Information and Computer Science with a minor in Technical and Business Communication in 1990 and an MS in Information and Computer Science in 1992 (also from Georgia Tech). After one year at the University of South Florida, Dr. Antón joined the computer science department at NC State. She was awarded an NSF CAREER Award in 2000, named a CRA Digital Government Fellow in 2002, nominated and selected for the 2004-2005 IDA/DARPA Defense Science Study Group, and received the CSO (Chief Security Officer) Magazine "Woman of Influence in the Public Sector" award at the 2005 Executive Women's Forum. She is associate editor of *IEEE Transactions on Software Engineering*, the cognitive issues area editor for the *Requirements Engineering Journal*, and a member of the International Board of Referees for *Computers & Security*. She is a member of the International Association of Privacy Professionals, a senior member of the IEEE as well as a member of the ACM U.S. Public Policy Committee's Executive Committee. Antón currently serves on several boards: the NSF Computer & Information Science & Engineering Directorate Advisory Council, the Computing Research Association's Board of Directors, the CRA-W Board, an Intel Corp. Advisory Board, the Department of Homeland Security Data Privacy and Integrity Advisory Committee, the Berkeley TRUST Center Distinguished Advisory Board, and the Georgia Tech Alumni Association Board of Trustees. She is a former member of the Microsoft Research University Relations Faculty Advisory Board and the Georgia Tech Advisory Board (GTAB). Dr. Antón is director of ThePrivacyPlace.Org (http://theprivacyplace.org), and co-director of the NC State Electronic Commerce Studio. Her URL is: http://www.csc.ncsu.edu/faculty/anton.

About USACM

USACM is the U.S. Public Policy Committee of the Association for Computing Machinery (ACM). USACM members include leading computer scientists, engineers, and other professionals from industry, academia, and government. (http://www.acm.org/usacm)

About ACM

ACM, the Association for Computing Machinery (http://www.acm.org), is an educational and scientific society uniting the world's computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

Appendix B – Understanding Identity and Identification

USACM

The Public Policy Committee of ACM

Understanding Identity and Identification

Professionals who work with issues of security and control use some terms to precisely describe access to resources and naming. These same terms have usage in general language, but the words frequently are used imprecisely and even misleadingly. When describing how security in information systems operate, and when formulating regulations or laws, it is important that these terms are understood and used precisely.

The purpose of this short document is to describe these important terms for readers who are not familiar with the more formal definitions. These related terms are *identification*, *authentication*, and *authorization*. Related concepts include *uniqueness* and *biometrics*.

Terms

Identification is associating a distinguishing label (*identifier*) with something within a specific group or context. You can identify someone by getting both their label and the context of that label. An ID card can provide both the name (e.g. "John Smith") and the context (e.g., "licensed driver"). Identification can also occur by providing only the context or group name, such as identifying oneself as a police officer, a student, a graduate of West Point, or a member of Congress by wearing an appropriate badge, uniform, or class ring. The reliability of an identification depends on the confidence that the distinguishing label and context actually apply to the individual in question.

Note that even when identification is reliable – and it often is not – it does not imply anything beyond being able to distinguish among items or people. Identification can be used to determine if someone is a member of a group or not, or among members of the group. If someone were to identify herself as "Snow White," that is an identification if she uses it consistently. In the context of a Halloween party or an Internet chat room, that may be a logical label to adopt.

A key concept is that identification does not need to be a standard name. It can be a nickname, a login, or a simple description, such as "I am the tallest one here" or "I am the one with red hair." Those are means to distinguish one person from another in a particular group context.

People are most often identified in social situations by their names. In the United States, these names are usually composed of a first (given) name, one or more middle names (usually), and a last (family) name. In other countries, names may be a single word, or

everyone may have a common family or middle name.

Uniqueness is when multiple items do not have the same identifier. Human names are seldom unique across a large enough population. For instance, there are many, many people named "John Smith" in the USA. If we also consider ancestors, then there may be even more individuals who have been associated with the same identifier (name).

We can further qualify an identifier to make it more specific and less likely to be a duplicate of another identifier. For instance, someone could be "John Smith who was born April 1, 1952 in Boise and whose mother was named Matilda." However, we cannot always be certain this is unique, and it is unwieldy to use in formal documents. Thus, we commonly use an artificial identifier that is generated and assigned in a manner that ensures that it is unique within context. For instance, Social Security numbers are supposed to be assigned without reuse, making them theoretically unique. Other identifiers (e.g., driver's license numbers) are similarly generated to provide uniqueness.

Authentication is the process of verifying – to some desired level of confidence – that a claimed identifier is valid and actually associated with a particular item or person. Often, this validation is performed by one or more persons inspecting the identification and authenticator(s). The authenticators can also be examined by some technical means, such as a login program or a badge reader connected to a computer.

Authenticators of people are typically some combination of "something known," "something possessed," and "something about (structural)" the person. These items have been previously registered with the persons or organizations performing the authentication. Additional factors can also be used, such as physical location, recognition by human or canine guards, and so on.

- *Something known* is a secret or a fact that is unlikely to be known to an impostor. Passwords, when properly chosen and protected, are this form of authenticator. In many old combat movies, the spy is exposed because he doesn't know which team won the World Series the previous year – this is another form of "something known" as a group authenticator. Many companies use items such as "mother's maiden name," "birth date" or "social security number" as authenticators, but this is bad practice as those items are often easily discovered facts: Many of these items are public information as a matter of law or custom.
- *Something possessed* is a distinguishable token or a key that matches a counterpart. A license issued by a government agency is a form of token. Another example from an old movie is the dollar bill or playing card that is ripped raggedly in half – the two halves are kept and joined together to *mutually authenticate* two parties.
- *Something about* (structure) the object or person being authenticated. We can examine something physical about the person we wish to identify, such as a fingerprint, or the pattern of blood vessels inside the eye. If the comparison of a

person's distinguished characteristic is automated, then it is known as a *biometric*. A current location may also be used for authentication, such as GPS coordinates, telephone caller-id or computer network address.

Using a combination of authenticators is known as *multi-factor authentication*.

Authorization is the granting of rights (verb) or the grant itself (noun). Generally, one authorizes an authenticated party. *Permission* is used by some people as a synonym for authorization.

An example

Consider a scenario involving a person who wishes to enter a guarded building. When the person approaches the building to enter, a guard stops him to verify that he can enter. The person produces an *identification* card (something possessed) issued by a trusted authority (the context). The guard compares the picture on the ID with the face of the person, and causes him to put his fingers on a scanner (a biometric). These checks confirm that the person is the one identified by the card. She has been instructed that anyone with a valid blue card is allowed to enter, but without a cell phone, so she allows the person to pass after determining that he does not have a cell phone.

Note that this is use of multi-factor authentication, and the identification is based on group membership ("people with a valid blue badge") – no specific name or ID number is required. Permission to enter is the authorization involved. A further element of access control that is not based on identity or authentication is also involved: there is no authorization to carry a cell phone in.

There are many potential weaknesses in this system as described. The system can be redesigned to prevent the weaknesses, but defensive measures may be too expensive or cumbersome to be worth the effort given both the likelihood of the threats occurring and the value of what is behind the door. Examples of weaknesses include:

- The picture on the card may be old and the guard makes a false negative authentication: she refuses to allow the authorized person to pass.
- The guard may be overpowered or bribed so that unauthorized people enter.
- The card has been altered from a valid card — the color has been changed, or the original holder's photograph and fingerprints have been replaced by this impostor.
- The cards are made to published standards without adequate safeguards: this is a forged card made by a well-informed and sophisticated attacker.
- The attacker has stolen the card, disguised himself as the cardholder, and donned fingerprint caps that fool the scanning machinery.

- The guard is unable to recognize a disguised cell phone.
- Someone pretending to be a law enforcement officer, in uniform, orders the guard to let him pass or he will arrest her for obstructing justice. She complies.
- If too many people arrive in a short time, the guard may not be able to process them in a timely fashion, and someone is either denied access incorrectly or slips in unnoticed.
- The guard may fall ill and leave her post, leaving the door locked or unlocked for subsequent visitors.
- A first-time visitor has no way of knowing that this is really a legitimate guard and the right door!

Additional Notes

1. As illustrated by the last point in the previous example, the problem of authentication is bidirectional — all parties in the transaction need some level of assurance that they know the identities of the other parties. This is one reason why *phishing* succeeds: the customers enter their authenticating information, but the other party (the purported merchant) is not strongly authenticated to the customer.

2. It is possible to have authentication and authorization without specific identification. For instance, producing an *authentic* $20 bill provides authorization to make a purchase for something up to $20 in cost. It is not a requirement to *identify* the purchaser beyond being a member of the group who has cash.

3. Knowing precise, authentic identity **does not disclose intent!** Knowing the name of everyone who enters a building or boards a plane does not mean that they will be well-behaved. Mohamed Atta's Florida driver's license and picture were legitimate and examined when he passed through airport security on 9/11/2001. Most identification checks instituted in the wake of 9/11 perform at most a weak security function because there is poor (or no) authentication, and even when the identity is known it does not prove anything about intent.

4. Social security numbers are not supposed to be reused. However, numerous recorded cases of SSN duplication make the use of these numbers as unique IDs problematic.

5. Most biometrics have been developed and tested for authentication of a claimed identity, not for performing the identification itself; fingerprints are a notable exception. Insufficient experience has been gained with both physical features and biometrics to know error rates over large populations. By example, given the data that John Smith is 6'1" tall, has brown hair and green eyes, we can determine with some confidence whether a person in the room claiming to be John is actually John.

However, given that same information and a crowd of people in a football stadium, we cannot be certain that we can uniquely identify John if he is present. Almost certainly, we will also make many false positive identifications. The same problems may exist with automated biometrics such as measuring facial features or hand geometry.

6. We know that every potential biometric has deficiencies. Not everyone has valid fingerprints over their entire lives, twins and triplets have the same DNA, and so on. People with special interests in some technologies have made unsupported claims about the performance of certain biometrics.

7. Most organizations use weak authenticators. In part, this is because most people are poor at remembering items such as long passwords and multiple ID numbers. As noted, use of authenticators such as mother's maiden name, social security number, or other such items is poor practice because those items can be easily found for many people.

8. Every instance where identifiers and authenticators are to be used should be carefully analyzed to determine strengths and weaknesses. This includes the value of what is being protected, and the consequences of false positives (authenticating an incorrect identity) and false negatives (failing to authenticate a valid identity).

9. As noted, identification and authentication mechanisms depend on context. Any security protocol is only as strong as the weakest element.

Appendix C – Privacy Policy Recommendations

USACM

The Public Policy Committee of ACM

Policy Recommendations on Privacy

June 2006

BACKGROUND

Current computing technologies enable the collection, exchange, analysis, and use of personal information on a scale unprecedented in the history of civilization. These technologies, which are widely used by many types of organizations, allow for massive storage, aggregation, analysis, and dissemination of data. Advanced capabilities for surveillance and data matching/mining are being applied to everything from product marketing to national security.

Despite the intended benefits of using these technologies, there are also significant concerns about their potential for negative impact on personal privacy. Well-publicized instances of personal data exposures and misuse have demonstrated some of the challenges in the adequate protection of privacy. Personal data — including copies of video, audio, and other surveillance — needs to be collected, stored, and managed appropriately throughout every stage of its use by all involved parties. Protecting privacy, however, requires more than simply ensuring effective information security.

The U.S. Public Policy Committee of the Association for Computing Machinery (USACM) advocates a proactive approach to privacy policy by both government and private sector organizations. We urge public and private policy makers to embrace the following recommendations when developing systems that make use of personal information. These recommendations should also be central to any development of any legislation, regulations, international agreements, and internal policies that govern how personal information is stored and managed. Striking a balance between individual privacy rights and valid government and commercial needs is a complex task for technologists and policy makers, but one of vital importance. For this reason, USACM has developed the following recommendations on this important issue.

RECOMMENDATIONS

<u>MINIMIZATION</u>

1. Collect and use only the personal information that is strictly required for the purposes stated in the privacy policy.
2. Store information for only as long as it is needed for the stated purposes.
3. If the information is collected for statistical purposes, delete the personal information after the statistics have been calculated and verified.
4. Implement systematic mechanisms to evaluate, reduce, and destroy unneeded and stale personal information on a regular basis, rather than retaining it indefinitely.

5. Before deployment of new activities and technologies that might impact personal privacy, carefully evaluate them for their necessity, effectiveness, and proportionality: the least privacy-invasive alternatives should always be sought.

CONSENT

6. Unless legally exempt, require each individual's explicit, informed consent to collect or share his or her personal information (*opt-in*); or clearly provide a readily-accessible mechanism for individuals to cause prompt cessation of the sharing of their personal information, including when appropriate, the deletion of that information (*opt-out*). (NB: The advantages and disadvantages of these two approaches will depend on the particular application and relevant regulations.)
7. Whether opt-in or opt-out, require informed consent by the individual before using personal information for any purposes not stated in the privacy policy that was in force at the time of collection of that information.

OPENNESS

8. Whenever any personal information is collected, explicitly state the precise purpose for the collection and all the ways that the information might be used, including any plans to share it with other parties.
9. Be explicit about the default usage of information: whether it will only be used by explicit request (opt-in), or if it will be used until a request is made to discontinue that use (opt-out).
10. Explicitly state how long this information will be stored and used, consistent with the "Minimization" principle.
11. Make these privacy policy statements clear, concise, and conspicuous to those responsible for deciding whether and how to provide the data.
12. Avoid arbitrary, frequent, or undisclosed modification of these policy statements.
13. Communicate these policies to individuals whose data is being collected, unless legally exempted from doing so.

ACCESS

14. Establish and support an individual's right to inspect and make corrections to her or his stored personal information, unless legally exempted from doing so.
15. Provide mechanisms to allow individuals to determine with which parties their information has been shared, and for what purposes, unless legally exempted from doing so.
16. Provide clear, accessible details about how to contact someone appropriate to obtain additional information or to resolve problems relating to stored personal information.

ACCURACY

17. Ensure that personal information is sufficiently accurate and up-to-date for the intended purposes.
18. Ensure that all corrections are propagated in a timely manner to all parties that have received or supplied the inaccurate data.

SECURITY

19. Use appropriate physical, administrative, and technical measures to maintain all personal information securely and protect it against unauthorized and inappropriate access or modification.
20. Apply security measures to all potential storage and transmission of the data, including all electronic (portable storage, laptops, backup media), and physical (printouts, microfiche) copies.

ACCOUNTABILITY

21. Promote accountability for how personal information is collected, maintained, and shared.
22. Enforce adherence to privacy policies through such methods as audit logs, internal reviews, independent audits, and sanctions for policy violations.
23. Maintain *provenance* — information regarding the sources and history of personal data — for at least as long as the data itself is stored.
24. Ensure that the parties most able to mitigate potential privacy risks and privacy violation incidents are trained, authorized, equipped, and motivated to do so.

USACM does not accept the view that individual privacy must typically be sacrificed to achieve effective implementation of systems, nor do we accept that cost reduction is always a sufficient reason to reduce privacy protections. Computing options are available today for meeting many private sector and government needs while fully embracing the recommendations described above. These include the use of de-identified data, aggregated data, limited datasets, and narrowly defined and fully audited queries and searches. New technologies are being investigated and developed that can further protect privacy. USACM can assist policy-makers in identifying experts and applicable technologies.