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

# Social Security Administration's Role in Verifying Employment Eligibility

14 April 2011

Statement of
Ana I. Antón, Ph.D.

Professor
North Carolina State University

Director
CSC Policy and Compliance Initiative &
ThePrivacyPlace.Org

Vice Chair, USACM

## Introduction

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

I am a professor at North Carolina State University in the Department of Computer Science in the College of Engineering. In addition, I serve as Director of ThePrivacyPlace.Org, a privacy research center collaboration between NC State University and Purdue University. I also 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 position as well as that of the Association for Computing Machinery's (ACM) U.S. Public Policy Council (USACM), of which I serve as vice-chair. With over 100,000 members, the Association for Computing Machinery (ACM) is the world's largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. USACM serves as the focal point for ACM's interaction with U.S. government organizations, the computing community, and the U.S. public in all matters of U.S. public policy related to information technology.

The stakes are high for E-verify. This largely automated system—currently in pilot operation—may ultimately serve as the single most important factor in determining whether a person may be gainfully employed in the United States. As such, it must take into account complex issues around identity management, security, accuracy, and scalability, among others. These are not solely technology issues. Computing technologies are powerful and can play a role in employment verification, but even the most modern technologies have limits. Congress, the Executive Branch, and possibly the Judicial Branch must make decisions on risks and tradeoffs on complex policy issues. Should the E-Verify pilot system continue to be expanded, careful, balanced and informed consideration should guide both the technical architecture and policy decisions. This statement is intended to inform the committee on the computing community's perspective on these challenges. In particular, I wish to make three points on the key technology and policy issues:

**1. E-Verify must accurately identify and authenticate the individuals and employers authorized to use the system in a layered, trustworthy manner before it is widely deployed.** Although no authentication technology is perfect (including biometrics), effective approaches to identity management are layered and do not rely on one point of identification. Unauthorized accesses to the E-Verify databases would compromise the identities of anyone whose information it manages, including American citizens and permanent residents. The current pilot does not provide this level of accuracy.

**2. Proof of success with a pilot is required before extensively scaling any software system.** The E-Verify system should not be scaled up until certain weaknesses are eliminated and the pilot is objectively audited against established metrics for success. E-verify should not be extended to verify individuals' status for anything other than employment until after the system has been fully deployed and the impact and implications of any such extensions have been carefully considered.

**3. Complex systems (such as E-Verify) are fallible and often misused or repurposed in ways that violate sound principles of security and good software engineering.** Adequate, appropriate alternative mechanisms are crucial for handling unforeseen challenges and errors after the system is deployed. Even with initial pilot system success, scaling complex software systems may result in cost and schedule overruns, system breakdowns, intrusions and even obsolescence. Moreover, mission creep adds to the complexity of software systems, increasing the risk of the problems mentioned above. It also undermines the principle of data minimization as recommended in the USACM Privacy Recommendations (see Appendix C).

My testimony covers software engineering and security best practices that are relevant given our examination of the proposed expansions of the E-verify system. In this testimony, I describe several challenges for developing a system that securely verifies employment eligibility. Specifically, I discuss:

- alternative approaches to managing identity and authentication;
- plausible technical solutions for validating system pilots before proceeding to make it a permanent system; and
- objective, technical recommendations for this committee to consider as it moves forward with its efforts to verify employment eligibility in the United States.

## E-Verify Background

The E-Verify pilot system is designed to allow employers to determine whether an employee is eligible to work in the United States, using information reported in an employee's Form I-9 (Employment Eligibility Verification). Before the widespread use of digital technologies, the documents used to verify employment eligibility represented little threat of being a source of large-scale identity theft or fraud. However, now such documents are digitally scanned and incorporated into massive databases. If not properly managed the databases underlying E-verify could facilitate identity fraud and introduce significant risks.

Administered the by the Department of Homeland Security and the Social Security Administration, E-Verify is being used by over 238,000 employers, yielding 16 million queries[1] during 2010 the Fiscal Year. E-Verify is mandatory for some employers with federal contracts or subcontracts that contain the Federal Acquisition Regulation (FAR) E-Verify clause and employers in certain states.

[1] DHS E-Verify Web Page: http://www.dhs.gov/files/programs/gc_1185221678150.shtm

From a technical standpoint, difficulties in a pilot system's implementation provide reasons for concerns that would apply even more strongly if the pilot system's scale is widely expanded. These should be addressed before any further expansion of the pilot. Moreover, the significance of failures experienced with the pilot cannot be dismissed as acceptable and simply imposing additional mandatory training does not comprehensively address the problems. For example, a January 2010 audit report by the Inspector General[2] showed that the Social Security Administration itself failed to comply with the E-Verify Memorandum of Understanding (MoU) requirements. Specifically, the SSA: verified the employment eligibility of 26 existing employees because they had applied for new positions within the agency; erroneously verified the eligibility of 31 volunteers who were not employees; and verified the eligibility of at least 18 job applicants who were never hired—a clearly prohibited use. Moreover, 49% of SSA hires were not verified during the required 3 days prior to hire time period. Finally, the eligibility of 19% of the new SSA hires was never verified.

In December 2009, the Westat Corporation conducted an evaluation of the E-Verify system for the Department of Homeland Security[3]. Westat reported that 54 percent of the illegal immigrants checked through E-Verify were incorrectly deemed eligible to work because they are using stolen or borrowed identities[4]. This finding shows that the E-Verify pilot system is not able to detect identity theft and/or employer fraud.

The E-Verify pilot results from the SSA Inspector General and the Westat evaluation do not instill a sense of confidence that the pilot is ready to be promoted to a permanent larger-scaled system. Before scaling up, software engineering best practice requires a successful small-scale pilot and only when such success is achieved should one proceed to a larger-scale permanent system.

Scientific validation—evidence that a software system successfully meets pre-specified criteria with metrics—is critical before proceeding. The E-Verify pilot system includes policies and processes required for it to operate and perform as intended. However, flaws in business processes and factors external to the system can undermine an otherwise effective technology. As currently designed, there is no way for E-Verify to prevent any of the problems mentioned in the Inspector General audit report—improving, scaling and expanding the underlying technology will not solve the problems associated with erroneous verifications as system development continues.

A sense of urgency in our nation's efforts to protect its citizens can sometimes lead to taking shortcuts without proper validation and testing. Just last week, the Transportation Security Administration's (TSA) failure to scientifically validate their SPOT (Screening of Passengers by Observational Techniques) program before deployment was the subject of a hearing held by the

[2] The Social Security Administration's Implementation of the E-Verify Program for New Hires, Audit Report, The Office of the Inspector General, January 2010. http://www.ssa.gov/oig/ADOBEPDF/A-03-09-29154.pdf

[3] Westat Corporation, Findings of the E-Verify Program Evaluation (Rockville, MD), December 2009.

[4] Tim O'Coin, Study: E-Verify failure rate over 50%, WPRI.com Eyewitness News, February 25, 2010. http://www.wpri.com/dpp/news/local_news/providence-study-finds-everify-database-fails-to-catch-illegal-workers-over-50-percent-of-the-time.

House Science and Technology Committee's Subcommittee on Investigations and Oversight[5]. Validation and testing is especially important in high-value systems such as E-Verify. Compromises to these systems would likely result in massive identity fraud, which would be more damaging given the planned and proposed expansions to the E-Verify pilot system. Rushing deployment without fully addressing problems would also likely result in costly mistakes and overruns in implementation, which are not desirable at any time and especially not at a time of significant Federal budget deficits.

Several enhancements have been made to E-Verify since it was first introduced[6]. Since May of 2008, the Integrated Border Inspection System has provided real-time arrival and departure information for non-citizens. In February of 2009, DHS began sharing passport data and photographs with the Department of State (based on DoS records) as governed by a memorandum of understanding[7]. Both of these enhancements sought to reduce the number of mismatches in E-Verify. Such enhancements[8] are useful in that they are targeted and purposeful, serving to improve the system's ability to accurately verify individuals.

The new E-Verify self-check pilot allows workers to use the system to check their status without notifying employers or potential employers. Ensuring the system continues to only provides a simple "yes" or "no" response without revealing anything further is a step a step toward preserving the security of the system. However, we observe that there may be a potential for abusing self-check protection. For example, E-Verify could offer an unintended service to fraudsters, allowing them to validate identity data before attempting identity theft. The self-check pilot is available in six states (Arizona, Idaho, Colorado, Mississippi, Virginia, and the District of Columbia).

USACM reviewed the self-check pilot system[9] and noted that the system requested information that can easily be obtained via public records (e.g. county tax records) by individuals other than the holder of a given SSN[10]. Attention should be paid to whether the "what you know" questions for the E-Verify self-check pilot" are well-designed, meaning usable to the individual wishing to check their records, but unusable to outsiders. Our concern here is about the information that is being requested because it is not sufficient for proper authentication. The current selection of questions about the year in which you purchased your home, how much it cost and the age range

[5] Subcommittee Examines Behavioral Science Used by TSA to Screen Potential Security Risks, April 6, 2011. http://science.house.gov/press-release/subcommittee-examines-behavioral-science-used-tsa-screen-potential-security-risks

[6] "Priorities Enforcing Immigration Law," http://www.uscis.gov/portal/site/ uscis/ menuitem.5af9bb95919f35e66f614176543f6d1a/?vgnextoid=d3ace7c336c60210VgnVCM1000004718190aRCRD&vgnextchannel=8a2791daff2df110VgnVCM1000004718190aRCRD

[7] This memorandum of understanding was signed in December of 2008.

[8] Additional planned enhancements to E-Verify include: incorporation of Student and Exchange Visitors Information System (SEVIS) data, integration of DMV photographs (to date, no state has agreed to add its driver's license data to E-Verify), and allowing citizens to lock/unlock their SSNs for E-Verify purposes.

[9] Given that non-property owners are given a different set of questions, our tests can only be considered illustrative rather than comprehensive.

[10] In our experience, the self-check system requires an individual to submit his or her name, address, SSN and date of birth to access the system—information that is easily available to individuals wishing to verify someone else's employment eligibility. The secret questions cannot truly be considered "secret" given that the answers to these questions are available via public records: home addresses are available via whitepages.com; age range is available via whitepages.com; county or city in which one resides is available via Google maps; price paid for a home is available via local county tax records websites.

of the head of household, does not instill much confidence in this regard nor does the use of a "uscis.gov" URL for the "non-DHS, independent assurance service that uses non-governmental information to generate questions." Ultimately, we want citizens to be able to do their own self-checks; however, we must consider whether there are risks associated with granting unauthorized individuals access to the system or with allowing fraudsters to check the information they've stolen in an attempt to determine if they can use the information to craft a new, fraudulent identity.

## Mission Creep

Mission creep—also called repurposing or piggybacking—in software engineering refers to efforts to expand a system beyond its original goals after initial success. Mission creep introduces substantial risks associated with cost and schedule overruns[11], system breakdowns, and intrusions as new applications are developed and "linked" to existing systems without proper validation or architecting, resulting in (for example) brittle and vulnerable databases. The ACM U.S. Public Policy Council has been unable to obtain E-Verify pilot's pre-defined metrics for success. If criteria for success with specified metrics and thresholds for success have not been defined, then software engineering best practice suggests that the system should not continue to be extended, enhanced, or authorized to become a permanent system. As of December 2010, USCIS and SSA had not yet "established a written service-level agreement that describes acceptable and unacceptable SSA service levels required to support the E-Verify program"[12]. Defining these requirements is critical for establishing success criteria for the E-Verify program before scaling the system up.

Given the currently planned enhancements to E-Verify as well as the proposed legislation to expand the usage of E-Verify[13], one can envision years of continual pressure to expand its mission. Linkages to other databases and applications, whether for authorizing home loans or for denying certain services to deadbeat parents, will place tremendous pressure on a system designed for one specific purpose. In his 2007 testimony to this same subcommittee[14], Peter Neumann referred to this practice as "piggybacking," noting that each time a system or database is piggybacked it increases the system's exposure as well as the danger that the data integrity will be compromised and/or data will be leaked. Moreover, when data integrity has already been compromised, i.e. there are errors in the original database, those errors will then be propagated to the piggybacking systems. Thus, the potential impact of errors on individuals is progressively increased.

[11] F.D. Davis and V. Venkatesh. "Toward pre-prototype user acceptance testing of new information systems: implications for software project management," *IEEE Transactions on Engineering Management*, 51(1), pp. 31 - 46, February 2004.

[12] Federal Agencies Have Taken Steps to Improve E-Verify, but Significant Challenges Remain, GAO Report #GAO-11-146, December 17, 2010.

[13] There are three specific bills that seek to expand E-Verify: (1) H.R.693: The E-Verify Modernization Act of 2011, (2) H.R.695: Legal Eligibility for Granting A Loan Act of 2011, and (3) H.R.282: To require Federal contractors to participate in the E-Verify Program for employment eligibility verification.

[14] Testimony of Peter G. Neumann on the Security and Privacy in the Employment Eligibility Verification System (EEVS) and Related Systems, House of Representatives Committee on Ways and Means Subcommittee on Social Security Thursday, June 7, 2007.

We will note that past experience with large systems IT procurement and engineering has shown that adding new missions to existing systems results in delays, errors, and cost overruns. This has been the experience with procurements for systems in the Department of Defense, IRS, FBI, FAA, and many other Federal agencies. This also can introduce new vulnerabilities. Thus, we recommend extreme caution in any expansion of the E- Verify system beyond its original design.

## Authentication and Access Control

In addition to mission creep, one must consider the risks associated with authentication and access control.

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).[15] 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.

Government systems that rely on the SSN as an identifier and authenticator are risky. 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. Authentication needs to be performed in a way that someone eavesdropping on a transaction cannot then masquerade as either the individual or the government service system for any operation. Moreover, the authentication should not center on questions whose answers are easily obtained by a fraudster via public records that are available online (e.g. property tax records). In this regard, the E-Verify self check implementation is troubling.

One form of identity management and access control that is being proposed is the use of biometric solutions. Given currently available technology, the idea of a tamper-proof identity card is a myth. No identification is completely tamper-proof or secure because perfect security is simply not possible. For example, an attacker could steal or counterfeit the ID, etc. Ultimately, security is about risk analysis. Thus, it is important to focus on risk-based approaches to improving identification, such as counterfeit-resistance.

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

## Biometrics

Biometric technologies have been proposed by some vendors as a method to more accurately identify individuals in a manner that cannot be forged. These technologies offer several benefits. In particular, physical attributes are extremely difficult to forge or fake; using numerous attributes provide a high likelihood of uniquely identifying an individual, and biometrics are difficult to forget or leave at home when a token, card or fob is not required. However, there are several distinct disadvantages: biometric readers are expensive and some have significant failure rates, biometrics are irreplaceable—once collected, the information can never be recovered without trusting the collector, biometrics protecting high value objects or systems pose a threat to the owner (a thief may be willing to cut off a thumb, for example), and physical attributes change over time (a hand print taken with a particular ring may not work on a day when the owner forgot to wear the ring, and fingerprints become less distinct as one ages or due to years of manual labor).

Generally, there are two approaches to biometric identification technologies: a distributed token approach and a centralized database approach.

*Distributed Token Approach*[16]. In this approach, subjects are given a card or a token similar to a key fob that has a biometric reader on it. The reader is provisioned by imprinting the subject's biometric into it in a secure fashion. Once imprinted, the token can be activated by the subject by re-scanning their biometric. At that point the biometric sends its identification code to a reader. There are several advantages to this approach: there is no central repository containing biometric data, tokens can be programmed to use a new identification code if the old one becomes invalid (thereby avoiding the 'irreplaceable biometric' problem), different biometrics can be used in the same system depending on the readers (e.g. one person can use her thumb, another can use his index finger, another could use a vein/artery pattern in her hand). Finally, this approach is inherently secure because it is built as a two-factor authentication system—you have to use something you are (your biometric) as well as something you have (your token/reader device). This is an expensive approach, however, because everyone must be given a token upon provisioning. Although this approach still requires a central database, the database stores identification codes rather than biometrics. Within the context of E-Verify, an advantage for the government is that it would not require a database of biometric identifiers to be maintained. In fact, even if the biometric card, token or fob is lost or stolen, no biometric data is recoverable because its contents are encrypted. This is a huge benefit to security and privacy. However, a disadvantage of this kind of biometric technology is that it would require all E-Verify enrolled employers to purchase a biometric reader or scanner, introducing its own risks such as hardware failure.

[16] In a Distributed Token Approach a new biometric ID is first provisioned by identifying and authenticating and individual—the individual then receives a biometric token such as a card or key fob, which is imprinted with the individual's biometric. The token captures and encrypts biometric markers (e.g. a thumb print) in much the same way as a password is automatically hashed upon entry. This is important because if the token is lost, then no biometric data is compromised. Once a biometric token is imprinted, it is tested against a standard identification/authentication/authorization process to ensure accuracy. If the biometric marker is a hand print or thumb print, then the token owner would swipe their hand or thumb across a reader and the scan is automatically captured, encrypted, and compared with the biometric stored on the device. If it matches, then the token identifies itself through a secured channel to a device reader.

*Centralized Database Approach.* In this approach, biometric information is stored by an organization or the government in a remotely accessible database that is used to perform verifications. Within the context of E-Verify, one might envision a database that contains a "white list" of individuals who are authorized to work. The US-VISIT system employs such a database-only approach, but it compares biometrics against a "black list" of terrorists and other bad actors[17]. This black list approach would be less intrusive from a privacy standpoint, but its feasibility would be questionable given the number of individuals who wish to work in the U.S. but are ineligible to do so. It is easy to envision pressure to design E-Verify so that it would be capable of the same sort of comparison. Our USACM privacy principles emphasize that the least privacy-invasive alternative should always be sought in the design of any system.

A centralized database approach is less expensive than the distributed token approach because biometric readers can be stationed at access points and it does not require giving the subject a card or token. However, this approach requires the government to be trusted to protect irreplaceable biometric data and to not misuse biometrics for purposes other than the one for which it was collected. In addition, biometric databases are high-value assets and targets for criminals seeking to construct an ID. Moreover, this is a single factor authentication approach and, thus, it is a single point of failure.

## Maintaining Complex Systems

Given that mistakes can have serious human impact, any laws or rules should be carefully crafted so as not to hurt innocent individuals—especially those who may be victims of identity theft. In addition, E-Verify will continue to be an attractive target for mission creep because it offers an attractive way for some groups to suggest as a mechanism to identify individuals—for now as eligible to legally work (and, if Congress allows it, to verify individuals as eligible for an increasing number of services, including home loans).

The GAO has noted that USCIS and SSA currently lack the ability to accurately estimate costs for E-Verify, thus there exists a significant risk of making poorly informed decisions and not securing necessary resources, leading to cost and schedule overruns and performance shortfalls[18]. Numerous complex software systems developed for large government programs have exceeded their budgets while producing sub-standard software. The FBI's Virtual Case File system was abandoned in early 2005 after over 700,000 lines of code were produced and $170 million dollars were spent because the system failed to meet fundamental requirements outlined years earlier by the FBI[19]. It was replaced with another software development project, called Sentinel, which was deemed by the Department of Justice to be two years behind schedule and $100

[17] Black lists in technology are intended to be complete, but it is impracticable to identify every person not eligible to work in the United States. However, an incomplete black list could still prevent known bad actors from repeatedly attempting to bypass the system.

[18] Federal Agencies Have Taken Steps to Improve E-Verify, but Significant Challenges Remain, GAO Report #GAO-11-146, December 17, 2010.

[19] http://www.justice.gov/oig/testimony/0502/final.pdf

million dollars over budget[20]. The Department of Justice is also concerned that the original requirements for Sentinel are now six years old and are likely to be outdated by advances in technology. Similar schedule and budgeting problems affected the modernization of software systems at the IRS[21] and the FAA[22]. We have no technical assurance that a software system as complex as E-Verify would be developed without similar budgeting and scheduling problems. Moreover, the December 2010 GAO Report on E-Verify notes that the pilot system remains vulnerable to identity theft and employer fraud.

Even large, highly technical, security-conscious companies that depend on their security practices for their very existence experience security violations. In January 2010, Google announced that it had several systems compromised by a cyber attack known as "Operation Aurora[23]." In addition, Adobe, Yahoo!, Symantec, and Morgan Stanley were also attacked. Last month, RSA, Inc.—the firm that invented the first public key encryption algorithm for both signing and encryption—had sensitive information related to their popular two-factor authentication product called SecurID stolen. These incidents demonstrate the kind of attacks that target significantly important system or high-value asset (such as the E-Verify database) and which will be inevitable over the course of time.

## Recommendations

Here, I present two sets of recommendations. The first set of recommendations are technical in nature and address best practices in moving forward with the E-Verify pilot system. The second addresses broader public policy considerations based on experience with large, public-facing software systems.

### Technical Recommendations on Best Practices for the E-Verify Pilot System

- The E-Verify pilot system should not be scaled up or extended to verify individuals for anything other than employment until weaknesses, such as those identified in the SSA Inspector General audit and the Westat Corporation evaluation, are eliminated and the pilot is objectively audited to verify pilot success. Moving forward without proper system validation and verification will inevitably lead to cost and schedule overruns, system breakdowns, intrusions and perhaps obsolescence.

- It is imperative that vulnerabilities be examined and risks addressed to protect the system as well as the identities of the individuals whose information is managed within it. E-Verify remains vulnerable to identity theft, employer fraud and may serve as a valuable tool for identity fraudsters.

[20] http://www.justice.gov/oig/reports/FBI/a1101.pdf
[21] http://news.cnet.com/IRS-trudges-on-with-aging-computers/2100-1028_3-6175657.html
[22] http://www.gao.gov/new.items/d09271.pdf
[23] http://www.mcafee.com/us/threat-center/operation-aurora.aspx

- Although it is tempting to resort to use of biometric technologies as a solution to the authentication problem posed by a system such as E-verify, it would be premature at this time. Until further testing and consideration is performed, the use of biometric methods should not be considered for the following reasons:
  - No single biometric technology is applicable to the entire population: not everyone has all their fingers, or irises, or other body parts that might be measured uniquely. DNA is present across all living humans, but is the same in identical twins and triplets, and is expensive and slow to analyze.
  - Most biometric measures may change with time. Even fingerprints may wear away from age, medication, or labor (e.g., bricklayers and some chemical workers).
  - No large-scale studies have been performed on populations as large as the U.S. to determine the rates of collision and accuracy of biometric identifiers, or of the accuracy of biometric measurement devices.
  - In the event of some future compromise of any large biometric database of U.S. citizens there would be no way to "reset" the biometrics to start over if a way was found to forge their use.
  - Biometric collection technologies are susceptible to privacy abuses[24].
  - Many people are uncomfortable with biometric information being collected or used, and perceive it as an invasion of privacy.

Additional Recommendations Based on Experience with Large Public-Facing Systems

- Careful consideration is critical before mandating use of E-verify because mandatory use would basically also mandate an increase in computer fraud, abuse, and identity theft. If E-verify were to be made mandatory for every employer, it would be a burden on small employers and/or a major security problem. It would require small employers to install Internet connectivity that they might not have, including Internet ISP subscriptions from some rural and remote areas where such service would be expensive. Furthermore, most small businesses do not have either the expertise or the resources to properly secure those systems against viruses, botnets, and intrusions. Thus, their systems would be at risk, and the information they would enter about prospective employees would be at risk of exposure for identity theft.

- There should be strong penalties for employers taking action on non-confirmation returns without informing applicants, providing them an opportunity to appeal and correct mistaken information in the records. Otherwise, the system may be used as an excuse for employment discrimination. Because the E-Verify system is certain to have errors, failures, and be subject to problems verifying some special cases (e.g., victims of identity theft), it is all too

[24] Shimon Modi and Eugene H. Spafford; Future Biometric Systems and Privacy; chapter in Privacy in America: Interdisciplinary Perspectives; edited by William Aspray and Philip Doty; Scarecrow Press, Inc.; 2011.

easy for it to be used as an excuse that “the computer said you aren’t eligible” to deny someone’s application or status without investigating complaints of error.

- Exceptions for cases of natural disaster or emergency should be built in if E-Verify is mandated. For example, if there is another Hurricane Katrina where all personal records and identification is lost on a wide scale, or if an individual family loses all their possessions in a fire or tornado, presenting appropriate ID may be impossible. Requirements should be waived or suspended when seeking new employment under such circumstances.

## Acknowledgments

I am particularly grateful to Aaron K. Massey, Cameron Wilson (ACM Director of Public Policy), David Bruggeman (ACM Public Policy Analyst), Eugene H. Spafford (USACM Chairman, and Professor at Purdue University), Stuart Shapiro, Jeremy Epstein, Jessica Young, Ramya Gopolan, Peter Swire, Daniel Dean, Lane Surreat, Alan Sheridan, Vineha V Bhat, Prachi Nandgaonkar, and many other members of ACM and USACM for their generous help in my preparing this testimony.

### About USACM

USACM is the U.S. Public Policy Council 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 www.acm.org, is the world's largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the computing 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 A – Biographical Information

Dr. Annie I. Anton is a Professor of Computer Science in the College of Engineering at the North Carolina State University and Director of the Computer Science Policy and Compliance Initiative. 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. 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. Anton joined the computer science department at NC State. From 2005-2006 she was a visiting faculty (sabbatical) scholar at Purdue University's CERIAS. In 2008 she chaired the NC State Public Policy Task Force and she is currently chairing the NC State University Reappointment, Promotion and Tenure Committee.

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. In 2006, she was honored with an award for "Most Influential Paper of ICRE 1996" at RE'06 for her 1996 paper entitled "Goal-Based Requirements Analysis". Her 1994 IEEE Software paper with co-authors Colin Potts and Kenji Takahashi was ranked the #10 most highly cited IEEE Software paper in its 25th Anniversary issue. She is a former associate editor of *IEEE Transactions on Software Engineering*, and former 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 member of Omicron Delta Kappa (OΔK) National Leadership Honor Society, a senior member of the IEEE as well as Vice Chair of the ACM U.S. Public Policy Council.

Anton currently serves on various boards: the U.S. DHS Data Privacy and Integrity Advisory Committee, the CRA Board of Directors, an Intel Advisory Board, the Future of Privacy Forum Advisory Board, the DARPA ISAT Study Group, and is corporate secretary for Trekking For Kids, Inc. She is a former member of the Microsoft Research University Relations Faculty Advisory Board, the CRA-W, the NSF Computer & Information Science & Engineering Directorate Advisory Council, the Distinguished External Advisory Board for the TRUST Research Center at U.C. Berkeley, the Advisory Board for the Electronic Privacy Information Center in Washington, DC, the Georgia Tech Advisory Board (GTAB), and Georgia Tech Alumni Association Board of Trustees. Dr. Anton is director of ThePrivacyPlace.Org (http://theprivacyplace.org). Her URL is: http://www.csc.ncsu.edu/ faculty/anton.

## Appendix B – 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.

usacm ACM US Public Policy Council

**Appendix C – Privacy Policy Recommendations**

# USACM Policy Recommendations on Privacy

## 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 Council 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.