Association for Computing Machinery

Advancing Computing as a Science & Profession

Testimony of Peter G. Neumann
Principal Scientist, Computer Science Lab
SRI International

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

For the Congress of the United States
House of Representatives Committee on Ways and Means
Subcommittee on Social Security
Thursday, June 7, 2007, 10am
B-318 Rayburn House Office Building

Abstract

Security and Privacy in the Employment Eligibility Verification System (EEVS) and Related Systems

This testimony addresses some of the potential pitfalls that should be considered when planning systems with extensive computer database applications containing personal information, such as the Employment Eligibility Verification Systems (EEVS). Many of these concerns are also applicable to related programs such as US-VISIT and REAL-ID and to peripheral systems that may depend on EEVS or result from interconnections among those other systems. Widespread problems have arisen in efforts to develop complex systems that must satisfy critical requirements for security and privacy; these problems are also considered. Furthermore, there is a pervasive tendency to overestimate the benefits of computer-related technologies as would-be solutions to societal problems. We should not expect easy technological answers to inherently difficult problems. People are almost always the weakest links, although in many cases the system design and implementation create further weak links. A deep awareness of the long-term problems is essential before adopting legislation that might promise to help in the short term.

1. Introduction

Thank you, Chairman McNulty and Ranking Member Johnson, for the opportunity to testify at today's hearing exploring issues related to proposed changes to the EEVS. I commend you for exploring the policy and technology issues associated with current proposals to expand and make this program mandatory. The computing community has often seen problems that resulted from policies established without careful consideration of the inherent limitations of technology. This can result in serious technical and social hurdles, and can lead to problems that are difficult to remediate once they have occurred, but that could have been prevented proactively. We hope that your efforts can help to avoid such difficulties.

As Principal Scientist in the Computer Science Laboratory at SRI International (formerly Stanford Research Institute), where I have been since 1971, and as someone with 54 years of experience related to computer and communication technologies, I have explored the intersection of technology and policy in numerous contexts, with a particular focus on system trustworthiness, security, and privacy issues. These areas are particularly relevant to the technology and policy nexus because privacy and equal treatment under law are fundamental rights; technology can at the same time help secure and also undermine those rights -- depending on the policies and practices for its use. Privacy and security are inextricably linked. One cannot ever guarantee complete privacy, but the difficulties are severely complicated by systems that are not adequately secure. Creating complex systems that are dependably trustworthy (secure, reliable, survivable in the face of many adversities, and so on) remains a grand challenge of computer science. As we review a proposed expansion to the EEVS, USACM sees a number of issues that should be explored, debated, and resolved before adopting this massive new system for identity verification.

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). ACM is a non-profit educational and scientific computing society of more than 80,000 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.)

A brief biographical paragraph is appended.

2. Issues of Specific Concern in the EEVS

The information transmitted to and stored in EEVS includes all of the primary personal identifiers in the U.S. As such, any compromise, leak, theft, destruction, or alteration of this data would have severe consequences to the individuals involved, including, but not limited to, identity theft and impersonation. It is thus essential that the system be

designed, constructed, and operated with the quality of protection that is essentially that required for classified national security information.

2.1. Transmission of Information

Any legislation requiring the transmission of personal information across the Internet should require secure transmission of this information. Employers and agencies participating in the program should be required to have strong encryption, strong authentication, or even elementary security (such as Secure Socket Layer, SSL) for transmissions to and from employers. Calling out such specific technologies and details would be inappropriate for statutory language; however, the legislation should include performance-based standards for security that limit the exposure of personal information and provide accountability for every step in handling and processing this information. This will make it clear to agencies that implement the system, and employers who use the system, that the security of personal information is as valued by policymakers as the reliability and timeliness of responses. In the case of EEVS and many other important systems, it is much more important to have continuing trust in the security and accuracy of the information rather than to get results in the shortest possible time.

We recommend that legislation require that the system be designed to protect the integrity and confidentiality of information, that an independent security review evaluation be conducted before the system is deployed, and periodically after deployment, and that the results of these evaluations be made public. The systems and their operation should be required to follow Fair Information Practices. See also USACM's recommendations for database design (http://www.acm.org/usacm/Issues/Privacy.htm).

We further recommend that the legislation require security breach notification: if administrators become aware of any breaches that could potentially affect personally identifiable information, then they must publish a disclosure and must notify all individuals who may be affected. Congress could model this after various state disclosure laws, such as one recently passed in California.

We also recommend that individuals be notified whenever someone accesses their records. The cost would be small, relative to other costs of the system: one letter or e-mail per job application.

2.2. Accountability for Access to Information

Accountability from the end user to the system administrator is vital in a computing system for ensuring the integrity of the system. If people are not held accountable for their actions, then policies intended to curb abuse will be undermined as users circumvent policies to make their jobs easier. One way of improving accountability in any computing system is by requiring strong user authentication and access controls coupled with thorough tamper-resistant and tamper-evident logging of all activity. In addition, all system accesses should log who accessed which records, and individuals whose

information is stored should be informed who has accessed their records. This would then allow concerned individuals to detect misfeasance and improper access to their records. Each employer should identify a compliance officer (distinct from EEVS users). The system should automatically detect unusual user behaviors (to the extent technically feasible) and report them to compliance officers.

Some strong controls are clearly needed to explicitly bind the access of a particular request to a specific authorized requestor acting in a specific role for a specific employer. The same controls should be applied to the operators of the system. Names, titles, and SSNs of authorized system users are not enough.

Access controls are also critical if individual employees are going to access the system to check their own information. Procedures and policy need to be in place to restrict employees' access to only their own information. The ability to check the accuracy of one's own information is very important. However, such accesses also need to be controlled and audited, at least as extensively as the accesses on behalf of an employer -- particularly to be able to identify systematic misuses.

2.3. Scalability

To date the system has functioned as a pilot program. The pilot has about 8,600 employers (June 2006 number) registered, with about half of those employers considered active users. This is out of about 5.6 million employers (as of 2002) that would eventually use the system once the law is fully implemented. Just because it seems to work for a small number of employers does not imply that it would work for all employers. The scalability of EEVS is a very serious architectural issue, because it will have to handle at least a thousand-fold increase in users, queries, transactions, and communications volumes. As a general rule, each time a system grows even ten times larger, serious new technical issues arise that were not previously significant.

At present, eight percent of confirmation requests cannot be handled immediately. This percentage needs to be reduced significantly as the number of employers increases. This would reduce the frustration with the system as well as the additional time required for manual confirmation for those records that could not be immediately verified. The additional human resources and associated costs necessary to handle this burden must be taken into account and included in budgets.

In general, it is risky to operate a system outside its intended design capacity and rely upon it to work under all circumstances, unless it has been carefully designed and implemented with scalability specifically in mind. Issues relating to inadequate scalability could completely compromise the effectiveness of the resulting system.

2.4. Accuracy of Information

The system has weaknesses about the accuracy of information presented to the system by an employee or employer as well as the accuracy of the underlying databases.

Speaking to the first kind of inaccuracies - fraudulent documents - the GAO has indicated that the Basic Pilot cannot effectively detect identity fraud. Proposals to add a digitized photograph to any employment authorization document would help make sure the employer could confirm that the photograph on the documents matched the employee presenting them. However, it is unclear how much this would reduce identity theft.

The inevitable cat-and-mouse game that always occurs in security (an ever upward escalating spiral in measures and countermeasures) is likely to occur between the security control and those seeking to commit fraud. As it becomes known that photo verification is a security feature, obtaining official documents under false pretenses will become more valuable. This could be done by bribing an insider or providing fraudulent documents to obtain the identification. The fraud is simply moved to a different part of the system. We also note that requiring REAL-ID, as envisioned by the DHS's rules for implementation of the REAL-ID system, will not solve the insider threat problem. This was pointed out in USACM's comments on the REAL-ID rulemaking. (See the "insider threats" heading in USACM's comments: http://www.acm.org/usacm/PDF/USACM_REAL_ID_Comments_FINAL.pdf)

Carefully developed standards for digital photographs are necessary -- much like those for driver's licenses -- although they will not be sufficient for the prevention and detection of forgeries.

Serious areas of concern also exist for the second kind of inaccuracies -- bad information in the underlying databases, delays in entering or revising information, and inconsistencies and name confusions among different databases. The Social Security database is known to have a high number of errors in name matches, as well as some duplicate numbers. For example, the Social Security Administration's Office of the Inspector General recently estimated that the SSA's 'Numident' file -- the data against which Basic Pilot checks worker information -- has an error rate of 4.1 percent. If each of 5.6 million employers made a query of a different potential applicant, that percentage suggests that on average more than 200,000 of them might get false responses.

The other databases the system will rely on will have similar issues. We certainly recognize and endorse the importance of provisions that allow individuals to check the correctness of information in the system that relates to them. However, a better defined process of correcting any erroneous information would be the necessary next step in improving the reliability of these databases, and the system as a whole. The risks of incorrect information are considerable, although establishing standards and procedures for accuracy to avoid those risks and to remediate errors and malicious misuse is an extremely difficult task. Numerous potential employees could be wrongly denied employment because of inaccurate records, if this problem is not addressed.

Risks of identity theft and privacy violations are also present -- for example, if unauthorized or surreptitious accesses, or even changes, can be made. Explicit provisions

are needed to protect employees and potential employees from adverse consequences of database and data entry errors.

Employers should also be held accountable for misuse of their blanket access privileges, such as using the data for running credit and insurance checks, engaging in blackmail, and other inappropriate purposes.

USACM encourages Congress to consider undesirable effects of false-positive and false-negative results. (A false positive is when a response indicates someone may be hired, only to be overturned later. A false negative would be when a response indicates someone has not been confirmed, only to be shown later to be incorrect.) Given the possibilities for error, identity theft, and system failure, employers should be protected from penalties when acting in good faith, and potential employees should be protected against discriminatory behavior. This is a policy issue rather than a technical issue, but directly arises from using an imperfect system as an arbiter.

It must be possible for authorized staff, as well as potential employees, to challenge incorrect EEVS data and determinations.

2.5. National ID System Concerns

Although there is no national ID card requirement attached to the EEVS, the connections to various databases are similar to the REAL-ID system currently proposed by DHS. If the EEVS does store query information or holds duplicates of information gleaned from the databases it interacts with, then it will have the appearance of a national identity system. As the existence of a national ID is not authorized by the proposed Senate immigration reform legislation, the Department will need to take care to avoid even the appearance of providing such documentation. The tradeoffs here are extremely complex, but are probably already being discussed in other testimony and other hearings.

2.6. Accessibility Issues

The potential lack of timely and highly available remote access to EEVS is another concern. Many small employers may not have Internet access or even computers that would allow them to have access. Examples might include small shop owners who want to hire clerks, and farmers who want a few hired hands. Furthermore, access via slow-speed dial-up connections is not likely to encourage consistent system use. Real-time confirmation of employability is less likely to occur consistently in such cases, and in cases of loss of computing or communication connectivity.

Perhaps even worse, poorly protected systems and poorly trained users will probably fall victim to ubiquitous security vulnerabilities and malicious software on the Internet. Many casual or novice computer system users could become unsuspecting victims of scams, phishing attacks, identity theft, and so on -- as a consequence of being forced to add computing and connectivity to support use of EEVS.

It is also a certainty that criminal elements will craft phishing e-mail appearing to originate from the Department of Homeland Security. This would include pointers (URLs) to what appear to be DHS websites with the DHS seal and apparent certificates that are essentially indistinguishable from the real websites. Unsuspecting users who visit these sites might then be victimized, resulting in significant financial losses and other serious consequences that typically result from identity thefts. Skilled identity thieves are likely to be able to scam the system itself more readily than authorized individuals can protect themselves or correct data errors.

A further problem is that many of the computer systems used to access EEVS may not have adequate security, and may have been compromised. Unfortunately, the security of EEVS itself may be subverted by the lack of security in other connected systems (which potentially implies the entire Internet).

For these reasons, despite its possible benefits, EEVS might actually make identity theft easier and at the same time make remediation and recovery more difficult.

3. Broader Concerns

The current state of the art in developing trustworthy systems that can satisfy critical requirements such as security, reliability, survivability, and guaranteed real-time performance is truly very poor. This is not a newly recognized problem, and was well documented in 1990 in a report, Bugs in the Program, by James Paul (Subcommittee on Investigations and Oversight of the U.S. House Committee on Science, Space, and Technology). Subsequently, I presented four testimonies (1997, 1999, 2000, and 2001) for various House committees -- each of which suggested that the overall situation had incrementally gotten worse. Of specific relevance to this testimony was my written testimony for the House Subcommittee on Social Security, The Social Security Administration: PEBES, Identity Theft, and Related Risks, on May 13, 1997 -- now more than 10 years ago. Similar conclusions appear in my testimonies for Senate committees (1996, 1997, 1998). (These testimonies are all online, with links from my website, http://www.csl.sri.com/neumann.)

Software development fiascos abound -- including many highly visible projects that have been late, over budget, or indeed abandoned after many years and large expenditures. My Illustrative Risks compendium index (http://www.csl.sri.com/neumann/illustrative.html) cites numerous examples such as the IRS and Air Traffic Control modernization programs and the FBI Virtual Case File, to cite just a few. See also the PITAC report, Cyber Security: A Crisis of Prioritization: http://www.nitrd.gov/pitac/reports/20050301_cybersecurity/cybersecurity.pdf.

Privacy problems are also manifold, and becoming increasingly complex as ubiquitous dependence on computerized databases increases. The extent to which computer systems and databases can enforce privacy policies is severely limited by the absence of meaningfully secure systems, and by the number of privacy violations occurring outside

of the confines of the computer systems. Correctness and timeliness of the data are also major concerns.

Several problems with identity management must be addressed. The existing infrastructure is riddled with security and reliability vulnerabilities, and is not sufficiently trustworthy. 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. Identities are typically subject to masquerading and spoofing. Name confusions such as alternative spellings and aliases cause major confusions. Authentication is often compromised by "social engineering" and other nontechnological bypasses. Authorization is typically inadequately fine-grained (and worse yet, often supposedly all-or-nothing, but bypassable). Blanket authorization should be avoided, observing the Principle of Least Privilege -- under which access authorizations should be restricted to just what is needed to accomplish that intended task rather than being overly broad.

It is also worth noting that there are cases where identities need to be masked. Examples include individuals protected under the Federal Witness Protection Program, individuals granted asylum from other countries and given new identities, undercover intelligence agents, undercover law-enforcement agents working criminal cases, and sky marshals. (Note that the Transportation Security Administration somehow lost the employee personnel records for 2003-2005.) All of these people need to have verifiable identities that stand up to any scrutiny, online or otherwise. Exposure of their real identities may result in their violent deaths, compromises of national security, and possible violence to their friends and families. Those individuals will likely need employment under their alternate identities, and it must be ensured that any system implemented for EEVS does not endanger their cover identities. The more that databases become cross-linked, the more difficult it becomes to prevent errors and leakage of such sensitive information. Furthermore, such linkages make these database systems higher-value targets for criminals.

The requirement of masking, aliasing, or otherwise providing alternative identities seems to create a fundamental conundrum: maintaining the accuracy of a critical database while simultaneously undermining its accuracy may impair the accuracy of other data in the process.

Past legislative efforts for improving accuracy and integrity of public databases have caused serious problems with the viability of other systems. For example, the Help America Vote Act mandated statewide-centralized voter registration databases that must verify the accuracy of records by matching them with drivers' license records. States such as California found that the data-matching requirements in practice led to high rejection rates in some counties, depending on how strictly the data was interpreted across databases. This had the effect of reducing, not improving, voter registration list accuracy, because legitimate voters were removed from the rolls because of address typos and name variants.

4. Conclusions

The problems identified in this testimony are fundamental in the context of EEVS-like systems. There are many risks. Essential concerns for system and data security, system and data integrity, and individual privacy must be anticipated from the beginning and reflected throughout design, implementation, and operation. Many potential slippery slopes must also be anticipated and avoided. Privacy requires a real commitment to creating realistic policies and enforcing them.

Experience has taught us that the design of information systems is subject to many pitfalls that can compromise their effectiveness. If EEVS is not appropriately implemented, it could -- like many past systems -- be subject to problems that include, but are not limited to the following:

- Difficulties in maintaining accuracy, correctness, and timeliness of the database
- Inconsistencies among widely distributed systems with distributed data entry
- A popular tendency to place excessive faith in the trustworthiness of the system's responses
- A common tendency to place excessive faith in the infallibility of identification, authentication, and access controls to ensure security and privacy
- The lack of scalability with respect to ever-growing enormous databases, massive numbers of authorized users, and consequent communication and access limitations
- The complexity of requirements imposed by noncompromisible auditing and accountability, both of which introduce further problems with respect to system security and integrity and with respect to data privacy
- The complexity of audit trails and notification of accesses to audit trails themselves
- The risks of exacerbated problems that result from mission creep -- as further applications tend to be linked to the originally intended uses, and as control of the above factors becomes less possible
- Similar risks related to feature creep, with or without any oversight and audit mechanisms.
- "Piggybacking" by other agencies -- e.g., law enforcement and DHS might want to place silent-hit warnings (as was considered in the late 1980s for the National Crime Information NCIC system) that would inform them who was seeking information for anyone who was under surveillance. Linkages with databases for deadbeat parents, student loan defaulters, and other applications might also be contemplated. Each such connection would expand the exposure of the system and the dangers of incorrect data and data leakage.

Congress should establish clear policies and required outcomes, rather than prescriptive or detailed technical processes or systems. The technical challenges to achieving the policies and outcomes should be fully documented in the Congressional Record of the legislation.

Considerably more focused research is needed on total-system approaches that address identity authentication, authorization, and data protection within the context of overall system architectures for security and privacy. (For example, some promising new developments enable the use of cryptography to enable certain queries to be answered without requiring decryption and release of excessive information in violation of the Principle of Least Privilege. These techniques appear to be significantly less subject to misuse, including insider misuse.) Such approaches may be more effective than trying to rely on biometric and other devices whose effectiveness may be compromised by technological or operational flaws in the systems in which they are placed and errors in human judgment. Finally, incentives are needed to ensure that research and innovative prototypes are relevant to the real-world problems and to ensure that these advances find their way into the development and operation of practical systems.

Although similar comments can be made about REAL-ID and any other national identification systems, all of these concerns are specifically relevant to systems such as EEVS.

We have not attempted to be complete here, but rather to focus on the main issues. There are many relevant reports of the Government Accountability Office, the National Research Council, and other sources that I hope you have already seen. Whereas USACM and I speak from a technical perspective, we recognize the political imperatives regarding immigration and employment. We urge the Congress to focus on creating the right incentives for operators and employers that maximize achievement of our immigration laws and each citizen's right to work while minimizing privacy invasion, ID theft, and criminal activity. In this effort, technology should be seen as a supporting block, not the keystone of the arch.

We look forward to any further questions that might arise from your reading of this written testimony or from my oral testimony.

Acknowledgments

I am particularly grateful to Cameron Wilson (ACM Director of Public Policy), David Bruggeman (USACM Public Policy Analyst), Eugene Spafford (USACM Chairman, and Professor at Purdue University), Jim Horning, and many other members of USACM for their generous help in my preparing this testimony on rather short notice.

Contact Information

Peter G. Neumann
SRI International, Computer Science Laboratory
Menlo Park CA 94025-3493
Neumann@CSL.sri.com
http://www.csl.sri.com/neumann

Personal Background Information

Peter G. Neumann (Neumann@CSL.sri.com) has doctorates from Harvard and Darmstadt. His first technical employment was working for the U.S. Navy in the summer of 1953. After 10 years at Bell Labs in Murray Hill, New Jersey, in the 1960s, during which he was heavily involved in the Multics development jointly with MIT and Honeywell, he has been in SRI's Computer Science Lab since September 1971. He is concerned with computer systems and networks, trustworthiness/dependability, high assurance, security, reliability, survivability, safety, and many risks-related issues such as voting-system integrity, crypto policy, social implications, and human needs including privacy. He moderates the ACM Risks Forum (comp.risks), edits CACM's monthly Inside Risks column, and is the Chairman of the ACM Committee on Computers and Public Policy (ACM-CCPP), which serves as a review board for RISKS and Inside Risks and is international in scope. He is also a member of USACM, which is independent of ACM-CCPP. He created ACM SIGSOFT's Software Engineering Notes in 1976, was its editor for 19 years, and still contributes the RISKS section. He has participated in four studies for the National Academies of Science: Multilevel Data Management Security (1982), Computers at Risks (1991), Cryptography's Role in Security the Information Society (1996), and Improving Cybersecurity for the 21st Century: Rationalizing the Agenda (2007). His book, Computer-Related Risks (Addison-Wesley and ACM Press, 1995), is still timely -- including many of the problems discussed above. He is a Fellow of the ACM, IEEE, and AAAS, and is also an SRI Fellow. He received the National Computer System Security Award in 2002 and the ACM SIGSAC Outstanding Contributions Award in 2005. He is a member of the U.S. Government Accountability Office Executive Council on Information Management and Technology, and the California Office of Privacy Protection advisory council. He has taught courses at Darmstadt, Stanford University, the University of California at Berkeley, and the University of Maryland. Neumann chairs the National Committee for Voting Integrity (http://www.votingintegrity.org). See his website (http://www.csl.sri.com/neumann) for prior testimonies for the U.S. Senate and House and for the California state Senate and Legislature, publications, bibliography, and further background.

Dr. Neumann is Principal Investigator for two SRI projects that are relevant to this testimony:

* Privacy-Preserving Credentials, one of several subcontracts from Dartmouth College, Assessable Identity and Privacy Protection, funded by the Department of Homeland Security, 2006-CS-001-000001-02, FCDA #97.001. The SRI project is part of a collaborative team project jointly with the University of Illinois at Urbana-Champaign, Cornell, Purdue, and Georgia Tech. The project is contributing some highly innovative cryptographic research and extensive system experience to the application of practical techniques for advanced identity management with demonstrations of applications that will include health care and finance but that have significant relevance to identity management generally.

* A Center for Correct, Usable, Reliable, Auditable and Transparent Elections (ACCURATE), NSF Grant number 0524111. ACCURATE is a collaborative effort with colleagues at Johns Hopkins, Rice, the University of California at Berkeley, Stanford, the University of Iowa, and SRI. It is examining techniques and approaches for voting systems, with particular emphasis on security, integrity, and privacy. SRI

Neumann contributes to the following DHS project:

* Cyber Security Research and Development Center (CSRDC), Department of Homeland Security, Science and Technology Directorate, DHS Contract HSHQDC-07-C-0006 to SRI International. CSRDC is providing extensive support for S&T Program Manager Douglas Maughan's R&D program. (http://www.csl.sri.com/projects/csrdc and http://www.cyber.st.dhs.gov)