Security and Privacy Risks in Voter Registration Databases (VRDBs)

Peter G. Neumann
Principal Scientist, SRI International, Computer Science Lab
333 Ravenswood Avenue, Menlo Park CA 94025-3493 USA
Tel 1-650-859-2375 Neumann@CSL.sri.com http://www.csl.sri.com/neumann

Washington DC, 30 November 2007

Workshop on Voter Registration Databases, organized by the National Academies' Computer Science and Telecommunications Board (CSTB) as part of a study sponsored by the U.S. Election Assistance Commission

[Caveat to the committee: Kristen Batch asked me to give a big-picture overview for this session. I will not be surprised if much of this position paper duplicates what you already know, or statements by other panelists. However, I hope the system oriented perspective will be useful.]

ABSTRACT

Databases containing personal identification information tend to engender enormous potential risks relating to computer security, system integrity, data errors, accountability, correctness, remediation of incorrect and inconsistent data, personal privacy, identity usurpation (including what is loosely called identity theft), and personal well-being. Past experience with the development and use of such databases is not encouraging.

Voter Registration Databases (VRDBs) need to anticipate all of those risks. Of particular concern to the ongoing CSTB VRDB study is the requirement to maintain accurate and up-to-date records of all eligible voters, correctly recording all necessary changes. VRDBs are especially vulnerable to accidental errors, willful misuse, data manipulation, denial-of-service attacks, and many other problems.

In this position paper, I consider many of these risks and examine some principles that might help overcome them. Responses are given to three explicitly asked questions. Several recommendations are discussed.

To put the risks and the problems they create into a broader perspective, Appendix 1 outlines some generally relevant difficulties that have been observed in complex database applications in other disciplines. Appendix 2 summarizes a set of principles for VRDBs [ACM2006] developed by the USACM committee of the Association for Computing Machinery chaired by Paula Hawthorn and Barbara Simons.

POTENTIAL RISKS IN VRDBs

It is important to realize that there are no simple procedures by which the necessary requirements can be easily satisfied. The depth of the problems that must be addressed is considerable. Thus, we begin with a consideration of the risks before turning to requirements and principles.

Elections require a total-system approach to system security, system integrity, data integrity, and voter privacy. They represent an end-to-end assurance problem in which every step in the process is today a potential weak link. Thus, the risks that must be considered are typically dispersed accordingly.

With respect to voting machines, considerable emphasis has been devoted in recent years to their integrity, reliability, and accuracy (e.g.,
ACCURATE: A Center for Correct, Usable, Reliable, Auditable and Transparent Elections, http://www.accurate-voting.org) or their lack thereof (e.g., the California Top-To-Bottom Review [Cal2007]. However, of particular concern here are the initial steps in the overall election process, involving voter registration (before or possibly during each
election) and voter authentication (whenever the identity and validity of each would-be voter may be challenged, resolved, or deferred through provisional ballots that are evaluated later).

Many voting system developments and election procedures have been established without adequate concerns for overall system security, integrity, and privacy. Many slippery slopes are not being anticipated, or are ignored altogether. Privacy issues seem to be sublimated, with no real commitment to creating and enforcing realistic policies.
Proactive establishment of system requirements and observance of principles for development and operation are both very important, but widely disregarded in practice.

There is a popular tendency to seriously overendow technology as a solution to human problems. This tendency is manifest in several ways, including but not limited to ignoring the following problems associated with voter registration and authentication:

* Difficulties in establishing and enforcing uniform statewide (not to mention national) standards for voter identification suitable for registration.

* Difficulties in maintaining accuracy and correctness of registration data, in view of name and address variations, marriages, inconsistencies among county rules, and so on, which are exacerbated when people move from one location to another (as was the case after Hurricane Katrina).

* Improper (e.g., fraudulent) use of Internet registration. (Arizona allows Internet registration with an electronic signature; many other states provide application forms and other registration support online [Ele2006].)

* Inconsistencies that can arise from data entry into statewide VRDBs by multiple authorities (registrars, clerks, system intruders).

* Inabilities of and irregularities in existing voter identification, authentication, and access control mechanisms. These are often compounded by inadequate oversight of security and privacy when voters actually attempt to vote.

* Difficulties in cross-referencing and coordinating records across jurisdictional boundaries, in the presence of administrative inconsistencies, communication impediments, and access limitations.

* Inability to detect people registered simultaneously in multiple precincts and even multiple states, which can be compounded by the presence of aliases and variant name formats. Remediation of errors

is itself would be a slippery slope. Note that public access to every state's VRDB could create new opportunities for companies such as ChoicePoint trying to identify duplicates and enable overzealous elimination of questionable voter registrations. (ChoicePoint claims to be the leader in identification and credential verification, but has been implicated in several suspicious activities. See Appendix 1.)

* Complexity of requirements imposed by noncompromisible auditing and accountability, which introduce further problems with respect to system security, data integrity, and data privacy.

* Risks of exacerbated problems that result from mission creep -- e.g., if further applications become linked to the originally intended uses, and as control of the above factors is not properly enforced.

* Misuses and improper reuses of the databases, e.g., for commercial, surveillance, vindictive, partisan, or other purposes.

* Inadequacies in establishing and enforcing VRDB system access controls, as well as difficulties in monitoring VRDB system use and being able to detect misuses and improper reuses.

* Inadequacies that become particularly difficult for persons with disabilities, or that in some ways disadvantage those persons.

* Serious risks for persons who must live with identities and other personal information that must be treated specially, as in intelligence agents, people under witness protection programs, and particularly those people in danger of bodily harm, for whom voting registry information made public could compromise their well-being. (According to an Electionline Briefing [Ele2006], at least 35 states currently make exceptions for certain persons. For example, California Elections Code 2166.5 and 2166.7 [CalCode] allows upon request the redaction of personal information for state employees, volunteers, providers and patients of reproductive health services, victims of domestic violence and stalking, and public safety officers. New Hampshire also has special procedures for victims of domestic violence and people under witness protection.)

In addition to all of these potential problems, the increased online availability of public-record VRDBs is likely to increase the likelihood of automated data mining, with opportunities for identity theft, unmonitored automated purges of voters, and many other risks.

Each of the above risks suggests the need for explicit requirements to address and ameliorate those risks.

Identification and authentication require serious study and open discussion. For example, requirements for a government-issued photo ID (as has been proposed in Georgia) open up the arguments for and against REAL-ID, plus concerns that this might further disenfranchise certain types of voters.

Correctness and timeliness of the VRDB data is essential. Oversight over data entry, subsequent alterations,

removal of entries, and possible misuse of the data all require noncompromisible mechanisms for accountability and auditing of basically every database operation -- as well as system alterations made by administrators or external vendor or other third-party upgrades.

Mission creep could work both ways. On one hand, desires for nationwide unique identifiers might suggest using Social Security Numbers or REAL-ID as voter identifiers. On the other hand, VRDBs could be used in conjunction with EEVS and law enforcement databases for purposes other than voter registration.

Statewide VRDBs are now mandatory under HAVA, as of 1 January 2006.
However, several states have not yet been able to complete procurement and deployment of HAVA-compliant registration systems, with varying problems experienced in Alabama, Illinois, Maine, New Jersey, New York, Wisconsin and Wyoming, and only interim use in California. North Dakota does not register voters, but is building a statewide database to record who has voted. Texas reportedly experienced many complaints about poor system performance of its Texas Election Administration Management system (TEAM) during the May 2007 primaries, and disenfranchisement of eligible voters who were (mistakenly? accidentally? intentionally?) removed from the database [Ele2007]. A Brennan Center report [Bre2006] provides detailed analyses of each state's responses to the HAVA VRDB requirement.

All of these concerns need to be addressed by any comprehensive approach to voter registration and voter authentication. In addition, large database systems are typically replete with numerous additional problems, such as development delays and overruns, project cancellations when the delivered system is clearly unable to meet expectations, serious consequences of erroneous data, relative ease of undesired insider and outsider manipulations, and so on.

Above all, the requirements must be stated prior to development and acquisition, and must be reasonably assured in any system before deployment and continually throughout operation.

A COMPARISON OF VRDBs AND EEVS

It is interesting to contrast the potential difficulties that may arise in voter registration databases with the problems already being experienced in attempts to develop the Electronic Employment Verification System (EEVS) -- which I have discussed in testimony for the Congress of the United States, House of Representatives Committee on Ways and Means, Subcommittee on Social Security [Neu2007]. The EEVS pilot study (also now being referred to as E-Verify [Epi2007]) reportedly has errors in the records of over 4% among the employees contained in the pilot study. An error rate that high in the eventual system would clearly be extremely disruptive.

The biggest difference between VRDBs and EEVS is that VRDBs contain data that is local to states or in some cases smaller jurisdictions, whereas EEVS and its successors will eventually contain data on every eligible employee in the entire country and will be accessed by every employer nationwide. Nevertheless, many of the lessons that must be learned in considerations of EEVS are also applicable to VRDBs. The mission-creep issue of using REAL-ID or some other national identifier is present in both. Accountability and auditability are essential in both.
Addressing the need to provide rapid human-oriented procedures to

rectify errors in the computer system data is clearly critical to both, particularly if the error rates in VRDBs are even only some fraction of what they are in the EEVS prototype.

Some further illustrative examples from other types of database applications are given in Appendix 1.

SPECIFIC QUESTIONS

Q1: PRIVACY CONCERNS

What privacy considerations need to be taken into account?

In this context, I consider primarily physical privacy and information privacy of stored data and possibly interactions with voters over electronic communications such as telephony and the Internet [Be+2007].

* Identification and authentication. New technologies often tend to have unforeseen privacy problems. For example, improperly protected smart cards may leak personal information. Furthermore, there is a tendency toward using such technologies for multiple purposes, such as social security, employment verification, passports, health care, and elections), rather than expecting people to carry many different unique identifiers. RFID chips are notoriously problematic. Real-time biometric scanners with wireless scanners can add to the privacy problems. Surreptitious surveillance may lead to clandestine tracking of individuals.

* Accuracy. Although the effects of erroneous VRDB data may not seem to be primarily a privacy issue, incorrect data misinterpreted by untrained officials could result in some unexpected consequences, such as denial of the right to vote, arrests, deportations, and panicked reactions of voters who feel unduly threatened.

* Identity usurpation. Identity theft (including masquerading) may seem to represent only a relatively small portion of the problems associated with misuse of identity information. However, its consequences are considerable. Thus, the risks relating to VRDBs must be considered, particularly theft of the right to vote -- such as organized matching of VRDB entries with recent obituaries before authorities catch up. (Dead people voting is apparently an old American tradition.)

* Other data misuse. Although past and present VRDB records might seem to present few privacy problems beyond identity usurpation, any mandated presence of REAL-ID or Social Security Numbers could lead to data mining and aggregation of information from the assorted state and other public and private records, particularly if some of those databases contain greater detail than others that would enable cross-correlation.

* Annoyance. Presence of phone numbers and e-mail addresses in any state VRDB systems could lead to automated calling and e-mailing, in addition to the postal mail access that is already commonplace.

Q2: PRINCIPLES

What principles guide your security decisions? How might these apply to voter registration databases?

An oversimplified set of election principles is given in my book, Computer-Related Risks [Neu1995]. These principles addressed system integrity, data integrity and reliability, voter authenticity, voter anonymity, vote confidentiality, operator authentication, and system accountability. Additional principles also addressed system disclosability (avoidance of proprietary code and data), system availability (despite accidental and malicious acts), system reliability, interface usability, documentation, and assurance. Shamos's Six Commandments are of course relevant in spirit. Many other principles are associated more broadly with the development of trustworthy systems, such the Saltzer-Schroeder principles and others discussed in [Neu2004], and these should be applied to election systems, control systems, and any other systems that must be trusted to perform correctly.

Based on past evidence and on the expected opportunities for future misuse, the VRDBs clearly represent significant weak links that can be exploited or accidentally invoked. Therefore, all of these principles are specifically relevant for the development and operation of VRDB systems. Anything that can compromise any weak link in the election process can potentially compromise entire elections, local, statewide, or in some cases even nationwide. In particular, a concise summary of a set of eight principles of the USACM committee is given in Appendix 2.

Q3: EVALUATION STANDARDS AND METRICS

What standard, adversarial test could be applied against each state's database? What would you include in such a test?

Reliance on standards and on testing is inherently an incomplete approach, and confronts several slippery slopes:

* There are no adequate existing standard adversarial tests that are applicable here.

* Static testing of individual state systems (hardware and software) is not very satisfactory. Many problems will arise only through human interactions, inactions, or typically unanticipated malicious behavior. Proactive design of the VRDB systems would be vastly preferable, although it is not likely to be found even among commonly used best practices.

* Standards tend to be lowest common denominators that minimally please election officials, system developers, and evaluators.

* The widespread use of proprietary software and proprietary evaluations that has prevailed in voting machines must not be perpetuated in RVDBs.

As a consequence of these and many other concerns relating to prevailing weaknesses in standards and in testing methodologies -- whether static or dynamic -- I am very wary of giving any credence to a strategy for testing, knowing that it would be only a very small tip of a very large iceberg.

CONCLUSIONS: WHAT CAN BE DONE, REALISTICALLY AND CONSTRUCTIVELY?

Considerably more focused research, development of trustworthy systems, and operational oversight are needed on total-system approaches to the overall problems of election integrity. The extent of the risks is generally much greater than recognized, and needs to be addressed explicitly. VRDBs are just one piece of the overall puzzle, although a very important part.

Considerable effort is needed relating to any identity management that might be used in association with VRDBs. Biometric-based identity cards and similar means of identification and authentication are generally not a panacea. They can often be misused, forged, or otherwise subverted. Integrity of VRDBs may be compromised by technological or operational flaws in untrustworthy systems on which the database software is implemented.

Use of best practices and principles is highly desirable, but never quite enough -- because too many opportunities exist for accidental errors and intentional misuses. More broadly, incentives are needed to ensure that research and development of these systems are relevant to the real-world needs of managing fair elections and to ensure that these systems take advantage of the best of what is known in the R&D communities. Developers, maintainers, and users of VRDBs should always tend toward caution in looking for simple solutions. The same statement also applies to legislators.

The recommended standards (as being embodied in the Election Assistance Commission's current revision of the earlier voluntary guidelines) need to encompass stringent evaluations such as those carried out by the California Secretary of State's Top-To-Bottom Review [Cal2007] that go far beyond the existing evaluation procedures. Truly independent evaluations are essential, although quality assurance and testing of a product in the absence of its use is never enough.

System security and database security need to be built in from the outset, rather than superficially relegated to procedural constraints.

In addition to rigorous authentication of people with access to VRDBs and supporting systems, differential access controls, auditing, monitoring, and accountability are all essential. Above all, built-in VRDB system accountability must employ systems designed to be trustworthy and noncompromisible with audit trails for all potentially relevant accesses to the databases (and their underlying operating systems).

Also essential is human oversight with respect to the quality and integrity of the information in each state VRDB, with respect to detecting and responding to serious misuses of the data, assurance that legitimate voters will be sent mandatory notifications of actions intended to be taken (e.g., their being disenfranchised on certain alleged grounds), implementation of effective procedures to ensure voters will have adequate opportunities to remediate errors, and so on. Oversight must also be accompanied with nonpolitical personnel, including Chief Privacy Officers and ombudspersons who can be truly independent of political pressures. Otherwise, the entire election process can be severely biased.

Above all, the likely risks outlined here must be explicitly anticipated and addressed.

ACKNOWLEDGMENT

Thanks to Joseph Lorenzo Hall for his comments. He noted that the exposure of protected information for people at high risk is often underrecognized. He also encouraged me to put a somewhat more positive spin on what might otherwise seem like a very gloomy view of the future -- which I hope I have been able to do in the final section.

REFERENCES

[ACM2006] USACM, Study of Accuracy, Privacy, Usability, Security, and Reliability Issues, February 2006.
http://usacm.acm.org/usacm/VRD

[Be+2007] Steven M. Bellovin, Matt Blaze, Whitfield Diffie, Susan Landau, Jennifer Rexford, Peter G. Neumann, Internal Surveillance, External Risks, Communications of the ACM, 50, 12, December 2007.
http://www.csl.sri.com/neumann/insiderisks07.html#12
This article is based on a paper by the named authors, ``Risking Communications Security: Potential Hazards of the ``Protect America Act'' (http://crypto.com/paa.pdf).

[Bre2006] Brennan Center for Justice, Making the List: Database Matching and Verification Processes for Voter Registration.
http://www.brennancenter.org/stack_detail.asp

?key=97&subkey=9185&init_key=9160 [split URL]

[Cal2007] California Secretary of State, Debra Bowen, Top-To-Bottom Review, August 2007. http://www.sos.ca.gov/elections/elections_vsr.htm

[CalCode] California Elections Code 2166.5 and 2166.7. http://www.leginfo.ca.gov/cgi-bin/displaycode ?section=elec&group=02001-03000&file=2150-2168 [split URL]

[Ele2006] Holding Form: Voter Registration 2006. ElectionLine. http://www.electionline.org/Portals/1/Publications/ERIPBrief13.final.pdf

[Ele2007] Statewide Voter Registration Database Status. ElectionLine. http://www.electionline.org/Default.aspx?tabid=288

[Epi2007] Spotlight on Surveillance: E-Verify System: DHS Changes Name, But Problems Remain for U.S. Workers, Electronic Privacy Information Center, July 2007. http://www.epic.org/privacy/surveillance/spotlight/0707/default.html

[Neu1995] Peter G. Neumann, Computer-Related Risks, Addison-Wesley, 1995. ISBN 0-201-55805-X.

[Neu2004] Peter G. Neumann, Principled Assuredly Trustworthy Composable Architectures, Technical report for DARPA, Computer Science Laboratory, SRI International, Menlo Park, California, December 2004. This report considers principles for the development and operation of systems and networks that must be trustworthy. http://www.csl.sri.com/neumann/chats4.html, .pdf, and .ps.

[Neu2007] Peter G. Neumann, Testimony for the House Subcommittee on Social Security, 7 June 2007. This testimony considers some of the risks likely to result from the development of the Electronic Employment Verification System (EEVS). http://www.csl.sri.com/neumann/house07.pdf

APPENDIX 1: SOME RELEVANT EXAMPLES OF PAST DATABASE PROBLEMS

[References to the ACM Risks Forum given below can be found online (http://www.risks.org).]

Numerous serious problems have arisen in the past in database systems related to governmental activities. These include projects that have been late, over budget, or even mothballed after many years and large expenditures of funding and manpower. Other problems include deployed systems that have had inordinate false positives and/or false negatives. A few of these are included here because of their potential relevance to VRDBs and particularly in the need to avoid the problems encountered

therein in any VRDBs. Additional examples can be found in my Illustrative Risks compendium index (http://www.csl.sri.com/neumann/illustrative.html). See also the PITAC report, Cyber Security: A Crisis of Prioritization, for further discussion of development difficulties: http://www.nitrd.gov/pitac/reports/20050301_cybersecurity/cybersecurity.pdf

* The Electronic Employment Verification System (EEVS), and its pilot system E-Verify, noted above.

* The FBI Virtual File System (VFS). An attempt to modernize and unify FBI database systems has floundered because of major development difficulties. (See RISKS-24.03, -24.38, and -24.62.)

* The California Statewide Automated Child Support System (SACSS), affectionately know as the Deadbeat Dads Database, experienced huge overruns and the development contract was eventually canceled. (See RISKS-19.47.)

* Student and Exchange Visitor Information System (SEVIS) rushed into production prior to testing; files mysteriously deleted or "misplaced"; unable to modify existing records (e.g., when someone has a baby); system slowness; random crashes; inadequate help support. (See RISKS-22.81.)

* U.S. Government healthcare database for disciplinary records and malpractice actions incomplete, inaccurate. (See RISKS-21.15.)

* Names of 4000 AIDS patients leaked to press in Pinellas County, FL (RISKS-18.48,53); former Health Dept employee and roommate charged (Reuters, 15 Feb 1997)

* News reports have long noted problems with the no-fly list (as in CAPPS II), which now contains over a third of a million names and has had over 50,000 people wrongly detained because of supposed name matches (for example, Senator Ted Kennedy and everyone named David Nelson).

* Numerous privacy breaches have affected Time Warner, Ameritrade, LexisNexis, Bank of America, Wells Fargo, and SAIC, to name just a few of the more publicized cases noted in the media, some of which are cited in the Illustrative Risks compendium index.

Use of one company's third-party databases has presented numerous privacy problems noted here, the first of which is particularly relevant to voting:

* Florida election erroneous disenfranchisement of thousands of voters traced to bogus ChoicePoint data; ChoicePoint blamed DBT, its data aggregator (RISKS-21.42).

* In 2005, ChoicePoint sent warning letters to more than 30,000 consumers whose detailed personal profiles had been obtained by identity thieves masquerading as legitimate business people, having established 50 fraudulent businesses. (ChoicePoint's databases reportedly contain over 19 billion public records.)

* Erroneous law-enforcement data from ChoicePoint: Privacy Foundation's Richard Smith discovered he had been dead since 1976, and had aliases with Texas convicts; Chicago woman misidentified as shoplifter and drug dealer, and fired (RISKS-21.42).

APPENDIX 2: SUMMARY OF THE USACM PRINCIPLES

The USACM study [ACM2006] is a well-reasoned analysis of many of the problems that may arise with VRDBs. The principles discussed there are outlined here:

1. The policies and practices of entire voting registration systems, including those that govern VRDBs, should be transparent both internally and externally.

2. Accountability should be apparent throughout each VRDB.

3. Audit trails should be employed throughout the VRDB.

4. Privacy values should be a fundamental part of the VRDB, not an afterthought.

5. Registration systems should have strong notification policies.

6. Election officials should rigorously test the usability, security and reliability of VRDBs while they are being designed and while they are in use.

7. Election officials should develop strategies for coping with potential Election Day failures of electronic registration databases.

8. Election officials should develop special procedures and protections to handle large-scale merges with and purges of the VRDB.

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, cryptography 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 represents an international rather than national 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.

Dr. Neumann has participated in four studies for the National Academies of Science: Multilevel Data Management Security (1982), Computers at Risk (1991), Cryptography's Role in Securing 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, and discusses many of the risks noted 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, the University of California at Berkeley, and the University of Maryland. He also chairs the National Committee for Voting Integrity (http://www.votingintegrity.org). See his website (http://www.csl.sri.com/neumann) for testimonies for the U.S. Senate and House and for the California state Senate and Legislature, papers, bibliography, and further background.

Dr. Neumann is the SRI Principal Investigator for A Center for Correct, Usable, Reliable, Auditable and Transparent Elections (ACCURATE), under NSF Grant number 0524111. ACCURATE is a collaborative effort that also includes colleagues at Johns Hopkins, Rice University, the University of California at Berkeley, Stanford University, and the University of Iowa. ACCURATE is examining techniques and approaches for voting systems, with particular emphasis on security, integrity, and privacy, and with much broader relevance to trustworthy systems with assurable auditing and accountability.

This workshop position paper was prepared with support from the SRI NSF ACCURATE Grant noted above.