More than a Matter of Trust: Security & Privacy in Voter Registration Records
Presented at the National Academies Committee on State Voter Registration Databases Workshop II

Bradley Malin, PhD  
(b.malin@vanderbilt.edu)
Assistant Professor of Biomedical Informatics, School of Medicine  
Assistant Research Professor of Computer Science, School of Engineering  
Vanderbilt University  
November 30, 2007

Questions

- What principles should guide security decisions? How might these apply to voter registration databases?
- What privacy considerations need to be taken into account, especially with the impact of combining and linking data?
- What standard, adversarial test could be applied against each state's database? What would you include in such a test?

Public Information Sharing

- Example: Washington
  - If you are a voter, your name, address, political jurisdiction, gender, date of birth, voting record, date of registration, and registration number are public information under state law. (RCW 29A.08.710)
  - This is public record by law and does not violate security or privacy, however …

Several Principles

- People are inherently curious.
- Account for the hacker, but don’t forget about the insider

- "Easier" Goals
  - Authenticate the users
  - Access Control
  - Logs and Audits
  - Encrypt data in transmission
  - Encrypt data at rest
  - Try to minimize data on mobile devices
  - Passwords, encrypt, etc.

- "Harder" Challenges
  - Define use cases
  - Specify workflows
  - Define roles
  - Develop Standards
  - Learn or specify baselines of systems use
Secure Data Sharing

- NAS Committee: Recommend how to evolve and maintain voter registration databases in a way that enables sharing of information with other states securely and accurately in fulfillment of the Help America Vote Act of 2002.
- Is data sharing centralized or distributed?
- What information is being shared?
  - Is it public information (names, dates of birth, etc.) or private information (Social Security Numbers)?

Private Information Sharing

- Is data sharing centralized or distributed?
- What information is being shared?
  - Is it public information (names, dates of birth, etc.) or private information (Social Security Numbers)?

Private Information Retrieval

What Data is Being Shared?

- Does voter registration, and other personal information, need to be compared in the clear?
- Computer science research has shown that comparison of encrypted records is possible.
  - Secure multiparty computation (e.g. Clifton; Wright; Nissim)
  - Need for standardized tools

Questions

- What privacy considerations need to be taken into account, especially with the impact of combining and linking data?
- Concerns are dependent on the realm
  - Information disclosed to the public is different than information used by the private data holders
- Threats exist primarily because we do not consider the availability of other databases

Privacy Concerns

- Privacy violations via inference: voter registration history
- Privacy violations via linkage: voter registration lists
- Untargeted Identification (UI): Identify any individual
  - Mass UI: Identify as many individuals as possible
- Targeted Identification (TI): Identify a specific individual
  - Mass TI: Identify as many pre-specified individuals as possible
- Ability to achieve attacks is dependent, in part, on the economics of data access
Privacy Violations By Linkage
(Sweeney 1997, 1998)

- Census and other government agencies have developed tools for analysis and protection of contingency tables for almost half a century

- Disclosure control
  - Statistical: suppress and/or permute voter histories with small cell counts (e.g. Fienberg; Duncan; Domingo-Ferrer)
  - Computational: generalize geocodings until cell counts are above minimal threshold (e.g. Sweeney)

- Principle

- 87% of the United States is RE-IDENTIFIABLE

- When it is revealed how Catoosa county voted in this election (aggregate results), then we uniquely link this voter to their vote.
Technology + Policy

- Must combine security and privacy risk evaluations with economic and policy deterrents
- Current regulations are insufficient: nothing to prevent the linkage of this information with other records.
- Legal protection from marketing, but does not prevent an adversary from linking a voter registration record to financial, health, and other types of information

Principles

- What standard, adversarial test could be applied against each state’s database? What would you include in such a test?
  - Develop a threat model
  - Model what other databases in the state are available?
  - Document what type of information the records contain
  - Be aware of who has access to these records
  - Rank risks by
    - Accessibility – How many people potentially have access to the data?
    - Sensitivity – What do the available records communicate about the individuals?
    - Cost – What are the economic barriers to access and linkage?