Vote Selling, Voter Anonymity, and Forensic Logging of Electronic Voting Machines

Sean Peisert   Matt Bishop   Alec Yasinsac

given at
HICSS'09
Waikoloa, HI
January 7, 2009

Meltdowns in Elections

- Most have been with regard to paper ballots, and not e-voting machines *so far.*
- A few high-profile problems (Sarasota Cnty., FL, 2007).
- Reasons to be concerned (possible vote-dropping in 34 states, 2008).
- Maybe more meltdowns we don't know about because the data is simply absent.
- This is bad: *intent of voters must be inviolate.*

E-Voting Machine Investigations

- Most investigations have been of software or machines, not election results:
  - California Top-to-Bottom Review (2007)
  - Florida ES&S iVotronic study (2007)
  - Ohio EVEREST (2008)
  - Operation BRAVO (2008)

What happens when something goes wrong with an election?

Post Mortem Procedures

- Collect evidence
  - Machines, printers, memory cards
  - VVPAT (if they exist)
  - System logs (if they exist)
  - Precincts' voting registers
- Look for discrepancies
  - Determine potential causes of discrepancies

VVPATs are not audit trails
(Yasinsac & Bishop, HICSS'08)

- If a VVPAT shows an undervote:
  - could be malfunction
  - could be voter choice
- If a VVPAT shows an over-vote:
  - probably malfunction, but where?
- If a VVPAT shows an equal balance:
  - implies that any problem did not involve dropping or adding votes (but could simply be mis-recording votes)

What is Forensic Analysis?

- Forensic analysis is the process of answering the questions:
  - How did an event take place?
  - What was the nature of the event?
  - What were the effects of the event?
- Forensic analysis applies to arbitrary events. This can include attacks, but is not limited to attacks (e.g., mistakes).

When We Need Forensics & Audit Logs

- Computer forensics in courts
- Recovering from an attack (including insiders)
- Compliance (HIPAA, SOx)
- Human resources cases
- Debugging or verifying correct results (e.g., electronic voting machines)
- Performance analysis
- Accounting

Forensic logging has been an essential element of validating security since at least 1980. Why isn't it done on e-voting machines?

- No real logging/auditing standards.
- No real consistent machine standards.
- No real legal guidance.
- In forensic auditing, accountability and traceability are key. That's exactly what *cannot* be done with voters.

Principles of Auditing Electronic Voting Machines

- Need to be able to count ballots
- Need to be able to determine if and how a machine failed.
- Cannot allow a voter to indicate to an auditor who they are (vote selling)
- Cannot allow an auditor to determine who a voter is (voter coercion)
- This leads to a direct conflict. So how do we balance this?
    - Add (benign) noise
    - Enforce (benign) regularity
    - Split data

Example of Conflict

- Voter: George, Auditor: John
- Scenario: George wants to sell his vote. John will pay for votes for Thomas.
- Forensic Audit Trail (FAT) records touches.
- John tells George to select James/Andrew/James/Thomas to identify himself in a FAT.
- This is a *covert channel.*

Example of Adding Noise

- If someone touches the screen in $x > \varepsilon$ places, can we assume communication, and add y additional touches without removing important information?
- If someone touches the screen in $x < \varepsilon$ places, we might suspect a mis-calibrated screen and/or undervotes.

Example of Enforcing Regularity

- If a voter casts a write-in vote, correct the spelling, capitalization, etc.., to the registered version.
- If a voter votes on initiatives in reverse order, have the logs reflect a forward order.

Example of Splitting Data

- If personally identifiable and/or data communicating a possible covert channel can be split from voting data, then two or more independent analysts can audit the data.
- E.g., separate ballot selections and transform multiple touches so that exact locations do not correspond to ballot

Audit trails are...

- It is is not well understood what forensic data is necessary, and there is no general solution to find that data.
- Data is often redundant, missing, vague, or misleading.
- Forensic analysis is worthless with bad data.
- We're wasting time, drawing bad conclusions, and making bad decisions.
- We need better data.
- A systematic approach to forensic logging gives better data and better analysis.

Erroneous and Missing Data

- The problem isn't just erroneous data...
  - we don't know have enough good data to identify/outvote the erroneous data
  - we don't know the assumptions, and so even accurate data may lead to erroneous conclusions
  - assumptions and accuracy need to be part of the model
- The problem isn't just missing data...
  - we don't know what's missing
  - we don't know what attacks we can't analyze without more/ different/better data
  - we don't know what attacks we can analyze with current data

What are the assumptions for e-voting and current forensic tools?

- Often that there's only one person who had access to the machine (what about sleepovers?).
- Often that the owner of the machine was in complete control (as opposed to malware or third-party virus scanners).
- *Probably a lot of other assumptions that we have no clue about...*

Current State of Forensic Tools

- Decent tools, but what problem do they solve?
  - file & filesystem analysis (Coroner's Toolkit, Sleuth Kit, EnCase, FTK)
  - syslog, tcpwrappers, Windows event logs
  - BSM
  - process accounting logs
  - IDS logs
  - packet sniffing

A Systematic Approach is Better

- Given system *S*, that records data *D*, what intrusions I_D can we understand with the data we have?
- Given intrusions I', what additional data $D_{I'}$ do we need to record to analyze those intrusions?
- Given an arbitrary system defined by certain specifications, what information must be logged to detect violations of those specifications?

Laocoön

- *Laocoön: A Model of Forensic Logging*
- Attack graphs of goals.
- Goals can be attacker goals (i.e., "targets") or defender goals (i.e., "security policies")
- Predicates represented by pre-conditions & post-conditions of events to accomplish goals.
- Method of translating those conditions into logging requirements.
- Logs are in a standardized and parseable format.
- Logged data can be at arbitrary levels of granularity.

Applying Security Policies

- Applying Laocoön to security policies guides where to place instrumentation and what to log.
- The logged data needs to be correlated with a unique path identifier.
- Branches of a graph unrelated to the attack can be automatically pruned.
- Defining policies and instrumenting systems can be hard on general-purpose computer systems.

Laocoön & E-Voting

- Good news:
  - Many violations of security policy on e-voting are easy to define precisely (e.g., changing or discarding cast votes)
  - Machines have (theoretically or ideally) limited modes of operation.

Possible Log Data

- Network traffic
- Insertion of new software
- Replacement of existing software
- System and library calls

Procedural Elements

- What about methods of bypassing the logging system?
- How tamperproof are logs?
- What about denial-of-service?
- What about human error?
- What about DREs vs. optical scanners?

Start with E-Voting Requirements

- Laws and requirements become security policies
- Security policies define attack graphs
- Attack graphs start with ultimate "goals"
- Attack graphs are translated into detailed specifications and implementations to guide logging
- Forensic data is used by an analyst.

Laocoön & Over-Voting

- Over-voting occurs when more candidates are selected than allowed in a given race.
- At some point, the value of a bit changes.
- What are the paths to that event?
  - Start with the entry to the system (e.g., touchscreen, supervisor screen, HW manipulation).
  - End at the data.
  - This places bounds on the intermediate steps.
  - Monitor those paths.

Summary and Status

- We need a means of verifying that votes have been recorded and tallied correctly.
- Forensics is an obvious solution.
- Current methods of forensic logging on e-voting machines is insufficient. VVPATs are insufficient.
- Detailed, systematic FAT is needed.
- FAT needs to be sanitized without removing important data.
- Some methods include adding noise, enforcing regularity, and splitting the data.

Going Forward

- Analyze covert channels and varying methods of sanitization on a specific machine
- Analyze means of integrating sanitization into e-voting system code base.
- Validation experiments (probably red teaming)

Thank you

- Questions?
- Sean Peisert
  - peisert@cs.ucdavis.edu
  - http://www.cs.ucdavis.edu/~peisert/