Alex Shankland

Electronic Voting Vulnerabilities and the Attribution Problem's Sibling

Abstract

Because most Electronic Voting machined are never connected to the internet, traditional means of incursion for the most part do not apply. However, the presence of other vulnerabilities, the age of the relevant software and hardware, and the extreme barriers of privacy involved in inspecting and investigating potential security breaches, mean not only that machines are insecure, but that identifying compromised machines themselves can be extremely difficult. This paper will outline the challenges to detecting election tampering at various stages in the process, and what noticeable irregularities each method will produce in the results, which can then be traced after the fact.

Introduction

One of the uncomfortable truths of the field of security is that, most of the time, people don't realized they've been breached until many days after the initial breach event.[1] Cybercriminals of any variety have little incentive to blast a pop-up window declaring "you've been hacked!" when they gain unauthorized access (with the caveat of ransomware complicating this slightly). Instead, discovering an attack, be it through malware, phishing, or stolen credentials, comes from the confluence of two factors: how often the relevant machine/data is used, and how much the presence of a hacker will noticeably change its behavior. In short, there have to be signs of a breach, and someone has to be paying attention to them. Electronic voting machines represent a combination of the worst possible aspects of both of these factors. They are rarely used (only during elections, with minimal inspections between them), and because of the strict requirement of voter anonymity, it is difficult to oversee them when they are in use. Because they are vulnerable, at varying points in their usage cycle, this poses a critical problem: how can one actually detect a hack when it occurs? Ultimately, most of the heavy lifting must be done through after-the-fact analysis of the results. Especially in the modern age where election forecast models, state-by-state exit polls, and highly accurate demographic data proliferate, any attempt to modify vote totals will leave irregularities which can be found.

[1] Verizon Data Data Breach Investigations Report, 2016, p. 10

To the Community:

I will be very honest: no matter how interesting it may seem at a glance, this kind of forensics is incredibly un-sexy. It involves very little writing of code and a whole lot of poring over excel spreadsheets. You will read these solutions and think, “well, there’s gotta be an easier way, I don’t even want to think about how annoying that would be to do.” That is exactly why I am writing this paper. If you ever are presented with a situation of figuring out if an electronic voting machine has been tampered with, you will not have direct access to the machine and a record of everything that happened to it. You will not be able to look over the software of the machine to find vulnerabilities, because to do so would be illegal. You will have to go at it in these oblique ways, and you will have to do the work of reconstruction, not merely protection.

The Problem as it Stands

It should be fairly obvious to anyone that voting machines can't be remotely taken over and accessed from your laptop, unlike other internet-connected machines such as personal computers, web servers, phones, and a surprisingly large number of cars, smoke detectors, watches, baby monitors, microwaves, etc. This is because voting machines, at no point in their life spans, are connected to the internet. They are air-gapped. However, this does not make them any more secure. According to the Brennan Center for Justice, this past election, some 43 states were using electronic voting machines more than 10 years old to service their shares of the 20% of the American public voting electronically.[2] These machines contain not only outdated software, but critical design flaws: in 2010, a pentester at the University of Michigan successfully converted a Sequoia AVC Edge paperless voting machine - a model currently in use in California, New Jersey, and Florida - into a Pac-Man game, without breaking any of the tamper-proof seals. It took him three afternoons.[3] How did he do this? He unscrewed the casing of the machine and swapped out its PMCIA card with one of his own design. Given the existence of the card, the full process would take only a matter of minutes, and could easily be accomplished in the space of a voting booth.

That was in 2010. More recently, researchers at Cylance managed to break into a newer version of the same model, and introduce software invisible to users and even inspectors, but which could not only change vote totals and precinct records, but also change the paper trail the machine left with each vote.[4] The solution they recommended was to put pieces of duct tape over

[2] Brennan Center for Justice, "Voting System Security and Accountability Risks", p. 2

[3] https://jhalderm.com/pacman/

[4] https://www.youtube.com/watch?v=Ss2fPvXBd7Y#action=share

the input slots and back panels of the machine. Duct tape is a very versatile mechanism and can solve many problems, but it is woefully inadequate here. The paperless voting machines in Pennsylvania, and these truly are paperless, they leave no paper trail with which to corroborate the vote totals, are stored in a locked warehouse near downtown Harrisburg.[5] While the address is not technically public and sharing it would be a misdemeanor, all government-owned buildings contain records of their appropriation in their state's budgets. Someone could then examine available photos of the only address in the warehouse district of Harrisburg, and observe that the security there is minimal. A good burglar with a sack full of memory cards could have flipped the tipping-point state of the 2016 election, and no one would have known.

Hello, there

Even securing the voting machines themselves, a difficult task at the best of times, is not enough: the paperless voting booths citizens see are only one part of the operation. The vote records from the precinct machines are then sent to aggregation machines at the county level, machines such as the AccuVote TS. Not only are these machines vulnerable, but they also create

[5] http://www.latimes.com/politics/la-na-pol-pennsylvania-voting-paperless-20161020-snap-story.html

a kind of network: vote totals are transferred to them in flash cartridges, essentially flash drives. These memory cards are moved from that aggregation machine to the voting machines, and back again, and are used interchangeably. This is exactly as bad as it sounds: in 2006, researchers at Princeton University created a voting machine virus which could infect a single voting machine of this time, then would spread to each aggregation machine used, and from those to every voting machine who got a cartridge inserted into those aggregation machines…[6]

The model they demonstrated this on has been largely, but not completely, retired. The same essential framework they exploited exists on nearly every single remote voting machine currently in use.[7]

[6] Feldman, Halderman, and Felten: "Security Analysis of the Diebold AccuVote-TS Voting Machine"

[7] https://www.verifiedvoting.org/which-voting-machines-can-be-hacked-through-the-internet/

Action Items

Given the clear problems with the infrastructure and software of electronic voting, including the intractability of the problem over a decade of its being known, it is apparent that merely patching these flaws is not enough. There must be a system in place to detect hacking as well. All of the exploits described above (with the exception of the pac-man) would be invisible to software engineers, who, for the most part, aren't allowed to examine the code of these voting machines. However, there do exist reliable ways of detecting ANY modifications to vote totals or other, related tampering. We will go into these further below.

Benford's law is a counterintuitive property of real world data: the leading digit of real world numbers tends to be low, following a specific probability distribution. Independent analysis of the county-by-county results from the 2000, 2004, 2008, 2012, and 2016 elections shows that this bears out for real world data, with probabilities approximately matching all the way down to the state level. Human beings, however, and particularly fixed percentages or human-influenced numbers such as those generated by voting machine malware, do not follow it: human beings use either randomness or their own intuition, both of which result in a uniform distribution. This method has already been used to identify possible fraud in the 2009 Iranian election, and is widely used in detecting fraud in the financial industry.[8] [9] Benford's Law analysis can be done with simple excel macros, but identifying where the violators are, especially if there are specific counties, can be difficult. For this reason, Benford's law alone is insufficient.

[8] https://www.newscientist.com/article/mg20227144.000-statistics-hint-at-fraud-in-iranian-election/

[9] http://www.journalofaccountancy.com/Issues/1999/May/nigrini

Demographic Indexing is a system that allows for far greater granularity. It exploits the human property that specific races and other identities frequently vote in varying and consistent proportions: black men will support different candidates in the aggregate than white women, and they will do so fairly uniformly, when accounting for other factors. For instance, geographically adjacent counties with similar demographic makeups will not differ by a significant amount. Independent analysis also bears this out: merely a naked-eye examination of county-by county results of the 2016 election will show that red counties and blue counties tend to be divided into distinct zones. Even with only state-level granularity in the exit polling, taking a random county's census data of demographics and reconstructing the voting pool based on the exit polls of its state provides accurate vote margins to within 1% accuracy. When considering how much voting must shift in order to flip the average state, this can be used not only to identify whether or not fraud took place, but where it happened.

Analysis using these two methods can corroborate reports of vote tampering with ease. What's more, they can be done after-the-fact by anyone. Demographic data on every single county in the country, as well as the county-by-county results of every election, are public record, and are for the most part available online. The analyses themselves are time-consuming, but not mathematically difficult. With an automated tool for Benford analysis and exit poll reconstruction, it would be possible to build a fraud detection system that could detect vote tampering on a national level, with county-level granularity, that could be run by one person and a powerful server.

Conclusion (and Supplemental Materials Index)

To conclude, the barriers to actually securing our electronic voter apparatus are unsolvable under our current framework. Instead, there is a more immediate problem: the fact that electronically-driven ballot tampering, of the types our system is vulnerable to, would not even necessarily be detected, despite the tools to do so having existed for as long as the machines have. There is a clear need for a detection net, capable of chewing through election result data and finding the subtle irregularities that would be present in the event of tampering.

As a demonstration of the following, I have attached the county-by-county breakdown of the 2016 election in the state of Texas. However, this dataset has been modified to simulate a large-scale election hack of the ballot virus format, effecting the entire state's infrastructure. In this election, Clinton won Texas, and with it, 270 electoral votes. To get a sense of what is involved, try finding which counties were effected using Benford's law, then,

Using CNN Exit polling here:

http://www.cnn.com/election/results/exit-polls/national/president

And County-level demographic data, which can be found by searching here:

http://www.census.gov/quickfacts/

confirm those guesses and by how much they were off. Good luck!

References

Verizon Data Data Breach Investigations Report, 2016

Brennan Center for Justice, “Voting System Security and Accountability Risks”

Feldman, Halderman, and Felten: “Security Analysis of the Diebold AccuVote-TS Voting Machine”

http://www.journalofaccountancy.com/Issues/1999/May/nigrini

https://www.newscientist.com/article/mg20227144.000-statistics-hint-at-fraud-in-iranian-election/