Testimony of Dr. Dan S. Wallach
Texas Senate Committee on State Affairs
October 15, 2008

Chairman Duncan, Vice Chair Williams, members of the committee, it's my pleasure to testify before you today about the security and reliability of electronic voting machines used in our state. I am an associate professor at Rice University in the Computer Science department. My research focus is on computer security and I have been examining electronic voting systems since 2001. I am also the Associate Director of the National Science Foundation's ACCURATE (A Center for Correct, Usable, Reliable, Auditable and Transparent Elections), a $7.5M research effort across six different institutions to improve our election systems. I have served as an expert witness in seven different cases concerning electronic voting, and I have also been part of several scientific analyses of electronic voting machines, most recently working for the Secretary of State of California as part of her "top to bottom" review, conducted last summer.

Present-day electronic voting systems have a variety of security flaws, many of which you've heard about. Of course, we can find problems with any voting system, but the present-day electronic systems enable fraud of a scale and simplicity previously unknown in the administration of elections. In the limited time available to me today, I'm going to discuss three kinds of failures in these systems and discuss steps that the state might take to address them.

Practical voting machine failures

First I would like to talk about real failures in real elections. These are cases where electronic voting systems have unquestionably failed. These are cases where the outcome of the election came under question. As you can imagine, the winner is always happy to win. The challenge with any voting system is to provide sufficient evidence to convince the loser that he or she lost.

Webb County, Texas. In March 2006, in Webb County's first ever election using its new ES&S iVotronic voting system, voters also had the option of voting with an optical scan ballot. In the primary judicial race between incumbent Manuel Flores and challenger Joe Lopez, Flores won on the paper ballots and lost on the electronic ballots. Out of roughly 50,000 votes cast, Lopez won with a margin of victory of roughly 100 votes: two tenths of a percent. I served as the expert witness for Flores.

In the limited time available, we were unable to find any evidence of fraud. What we did find was evidence of procedural errors on the part of the county elections administrator that raise serious doubts as to who should have won the election [W06]. For example, the "logic and accuracy" testing that they had performed consisted of casting one Democratic slate (always including Lopez for that particular race) and one Republican slate. I concluded that 26 such "test" votes for Lopez were included in the final election tally. Likewise, we found several machines that had been cleared on election day, causing an indeterminate number of votes to be lost. We also found votes recorded as occurring on days other than election day. We later determined these machines to have had their internal clocks set wrong by directly inspecting them. In the end, Flores conceded the race to Lopez, but the questions remain as to whether he won the race or not.

Sarasota, Florida. A more widely studied election failure, also involving the ES&S iVotronic, occurred in Sarasota County, Florida in the November 2006 general election, in the race for Congressional District-13 with Republican Vern Buchanan competing against Democrat Christine Jennings. Out of roughly 240,000 votes cast, there were more than 21,000 undervotes in this one race, and the margin of victory for Buchanan was 369 votes. I served as an expert witness for Jennings.

The cause of the undervotes is still disputed. One widely accepted interpretation is that poor design of the ballot layout caused these voters to simply not see the Congressional race and skip to the next race after it [HMH+08,AL08,FHHL08]. Another possibility is that machine malfunction may have contributed to the problem. (Many voters, while the election was ongoing, reported having problems with the machines.) Regardless, every expert who has examined the numbers agrees that, if the blank ballots were to be statistically reallocated based on how others voted, Jennings would have won the election. After a year of legal disputes, Jennings conceded the election and is now running again for the same seat.

Harris County, Texas. In November 2007, some Harris County voters were voting on a tax proposal. Apparently, 293 early voters never saw the question [B07]. As part of reconciling this issue, a Harris County election administrator used a feature of Hart InterCivic's Tally system called "Adjust Vote Totals" which does exactly what it sounds like. As it turns out, the way this feature works under the hood is that it simply replaces the totals in Tally's internal database. It leaves behind very little evidence that these "adjustments" were made. (For example, adjustments do not appear on final election reports.) Even without considering how this feature could be used in a fraudulent fashion, it's still amazingly error-prone. If you make a typo and don't catch it, it's very difficult to go back and undo any changes you might have made.

Human factors in voting systems

"To err is human, but to truly screw things up requires a computer," goes a famous saying. In recent years, researchers have begun conducting detailed, controlled human subject studies to learn how real voters might behave. These kinds of studies are incredibly valuable to our understanding of how these systems work and fail. For example, Herrnson and his team set up real voting machines in malls, nursing homes, and a variety of other places in Maryland, Michigan, and New York [HMH+08]. Their findings are fascinating. Whether on paper or with DRE systems, voters had consistently higher error rates when using straight ticket or write-in voting features [see p. 79, 85]. Voter-reported satisfaction varied, although they seemed to consistently dislike the Hart InterCivic eSlate, relative to its touch-screen competition [see p. 48-53]. This was also reflected in how accurately they were able to fill out their ballots [see p. 74].

In other studies, Byrne et al. [BGE07, EGB+08] found that paper ballots had consistently low error rates that were stable even across differences in age and education. Paper ballots yielded higher accuracy than DREs (in other cases, they seem to perform similarly; DREs are at best as good as paper, but not better). Despite this, voters preferred the DRE. In a subsequent study [S07], working with a DRE system we developed at Rice that can lie on its summary screen (you vote one candidate for president, but it either shows you another or simply doesn't show you the race at all), we discovered that *over 60% of test subjects did not notice when we manipulated the review screen*! Despite this, 95% of them reported that they felt the review screen was useful and they reliably preferred the DRE to the other methods.

In a nutshell, voters' subjective opinions of voting systems don't tell us much about how good these systems are at accurately and efficiently capturing voters' intent. Only through careful experimental studies, outside of real elections, can we ever learn what works and what fails.

Security vulnerabilities

I was first asked to testify about electronic voting systems before the Houston City Council in 2001. My opinion then, as now, is that computers are very easy to manipulate. Why should we believe that the election tallies are accurate? Efforts by others and myself have led to some serious analysis of these systems. In particular, I worked for the California Secretary of State last summer as part of her groundbreaking "top to bottom" review of electronic voting systems. I was on the team that examined the source code to Hart InterCivic's systems [CA-Hart07]. What we found was staggering.

Hart eSlate machines are connected to each other and a Judge Booth Controller (JBC) in a local network in the polling place. An attacker can plug into any eSlate and can send it a variety of commands. These include the ability to read and write to arbitrary memory addresses inside the eSlate. That means an attacker can extract all the votes from a machine and can replace them with anything else, all without detection. Similarly, an attacker can replace the software inside the machines with an arbitrarily malicious version. It's trivial to do this and still operate without triggering Hart's tamper detection mechanisms. Even worse, we found that a single corrupted eSlate machine, when it's brought back to the warehouse and connected to the "Tally" system (used for inventory control, among other things), it's possible to attack and corrupt the Tally system, which can then attack every subsequent eSlate. This is what we call a viral attack, and I cannot overstate the impact of this vulnerability. One attacker, corrupting one eSlate, in the current election can arrange for *every* eSlate to have corrupt software in subsequent elections. The only way an election official might be able to clean up or even detect a corrupt eSlate would be to open the case and replace the chips inside. Even if an attacker cannot manage to mount one of the attacks that I've described, it turns out that eSlates record votes in such a way that it's trivial to reconstruct the list of votes in the order they were cast. This could enable traditional voter bribery or coercion attacks.

The California study also considered Sequoia (not sold in Texas) and Premier/Diebold systems. The latter are also vulnerable to viral-style attacks, where regular election procedures can result in the spread of an infection from a single AccuVote-TS or TSx system to every other system in the county. A follow-on study conducted by the Secretary of State of Ohio [Everest07] confirmed all of our results and also found an equally staggering list of problems with the ES&S iVotronic, Unity, and other ES&S systems. In short, every electronic voting system used in Texas, both DREs and precinct-based optical scanners, are unacceptably vulnerable to very simple yet staggeringly effective security attacks.

Vendor response to these studies

Consider this statement from Hart InterCivic representative Peter Lichtenheld, testifying earlier this year before the Texas House Committee on Elections:

> Security reviews of the Hart system as tested in California, Colorado, and Ohio were conducted by people who were given unfettered access to code, equipment, tools and time and they had no threat model. While this may provide some information about system architecture in a way that casts light on questions of security, it should not be mistaken for a realistic approximation of what happens in an election environment. In a realistic election environment, the technology is enhanced by elections professionals and procedures, and those professionals safeguard equipment and passwords, and physical barriers are there to inhibit tampering. Additionally, jurisdiction ballot count, audit, and reconciliation processes safeguard against voter fraud.

Representatives from the other vendors seem to be saying similar things, so we can address this in general.

Did our work cast light on questions of security? Our work found a wide variety of flaws, most notably the possibility of "viral" attacks, where a single corrupted voting machine could spread that corruption, as part of regular processes and procedures, to every other voting system. In effect, one attacker, corrupting one machine, could arrange for every voting system in the county to be corrupt in the subsequent election. That's a big deal. (I separately wrote a paper delving further into this topic and how one should do risk analysis in light of these issues [W08].)

At this point, the scientific evidence is in, it's overwhelming, and it's indisputable. The current generation of DRE voting systems have a wide variety of dangerous security flaws. There's simply no justification for the vendors to be making excuses or otherwise downplaying the clear scientific consensus on the quality of their products.

Were we given unfettered access? The big difference between what we had and what an attacker might have is that we had some (but not nearly all) source code to the system. An attacker who arranged for some equipment to "fall off the back of a truck" would be able to extract all of the software, in binary form, and then would need to go through a tedious process of reverse engineering before reaching parity with the access we had. The lack of source code has demonstrably failed to do much to slow down attackers who find holes in other commercial software products. Debugging and decompilation tools are really quite sophisticated these days. All this means is that an attacker would need additional time to do the same work that we did.

Did we have a threat model? Absolutely! See chapter three of our report [CA-Hart07], conveniently titled "Threat Model." The different teams working on the top to bottom report collaborated together to draft this chapter. It talks about attackers' goals, levels of access, and different variations on how sophisticated an attacker might be. It is hard to accept that the vendors can get away with claiming that the reports did not have a threat model, when a simple check of the table of contents of the reports disproves their claim.

Was our work a "realistic approximation" of what happens in a real election? When the vendors call our work "unrealistic," they usually mean one of two things:

1. Real attackers couldn't discover these vulnerabilities.
2. The attackers can't be exploited in the real world.

Both of these arguments are wrong. In real elections, individual voting machines are not terribly well safeguarded. In a studio where I take swing dance lessons, I found a rack of eSlates two weeks after the election in which they were used. They were in their normal cases. There were no security seals. (I didn't touch them, but I did have a very good look around.) That's more than sufficient access for an attacker wanting to tamper with a voting machine. Likewise, Ed Felten has a series of blog posts [F08] about unguarded voting machines in Princeton.

Can an attacker learn enough about these machines to construct the attacks we described in our report? This sort of thing would need to be done in private, where a team of smart attackers could carefully reverse engineer the machine and piece together the attack. I'll estimate that it would take a group of four talented people, working full time, two to three months of effort to do it. Once. After that, you've got your evil attack software, ready to go, with only minutes of effort to boot a single eSlate, install the malicious software patch, and then it's off to the races. The attack would only need to be installed on a single eSlate per county in order to spread to every other eSlate. The election professionals and procedures would be helpless to prevent it. (Hart has a "hash code testing" mechanism that's meant to determine if an eSlate is running authentic software, but it's trivial to defeat. See issues 9 through 12 in our report [CA-Hart07].)

What about auditing, reconciliation, "logic and accuracy" testing, and other related procedures? Again, all easily defeated by a sophisticated attacker. Generally speaking, there are several different kinds of tests that DRE systems support. "Self-tests" are trivial for malicious software to detect, allowing the malicious software to either disable and fake the test results, or simply behave correctly. Most "logic and accuracy" tests boil down to casting a handful of votes for each candidate and then doing a tally. Malicious software might simply behave correctly until more than a handful of votes have been received. Likewise, malicious software might just look at the clock and behave correctly unless it's the proper election day. Parallel testing partly addresses this concern by pulling machines out of service and casting what appears to be completely normal votes on them while the real election is ongoing. Most Texas counties do not perform parallel testing and there's no state standard for how these should be done properly.

Auditing and reconciliation are all about comparing different records of the same event. If you've got a voter-verified paper audit trail (VVPAT) attachment to a DRE, then you could compare it with the electronic records. Texas has not yet certified any VVPAT printers, so those won't help here. (The

VVPAT printers sold by current DRE vendors have other problems, but that's a topic for another day.) The "redundant" memories in the DREs are all that you've got left to audit or reconcile. Our work shows how this redundancy is unhelpful against security threats; malicious code will simply modify all of the copies in synchrony.

Can the voting industry do better?

Voting system vendors and their trade organization tend to downplay the significance of third-party studies of their systems. Vendors typically point out that they have no evidence of attacks against their systems being attempted. Even if true, this doesn't discharge them of the responsibility to produce voting systems that do not have gaping security holes in their design. These vendors also like to point out how they are designed to meet the federal standards and the needs of their customers. That's certainly necessary, but it's demonstrably insufficient.

I wish I had confidence that the vendors could address these concerns. To date, we have the most long-term experience with Premier/Diebold, going back to a study on its security flaws that we first released in 2003 [KSRW04]. Five years later, they have clearly evolved their software, but haven't really improved their security in any meaningful way. This speaks as much to failures on the part of the vendor as to failures on the part of the federal and state certification processes. We simply cannot count on federal and state certification to ensure that our voting machines are secure. We cannot wait for the next versions of the vendors' software to be released and naïvely assume they will properly address all the shortcomings in the present versions.

If the vendors were serious about building stronger systems, they would be engaged in a public process of describing their future technologies and encouraging public and expert feedback. The vendors should be impressing us with their openness and clever designs, rather than hiding behind a standards and certification process that has demonstrably failed us all.

Public disclosure of vulnerabilities

The California teams did a huge amount of work, reading through these vendors' source code and cataloging their problems. They also produced "private" reports to the Secretary of State that contained much more specific information that would only aid an attacker or another security analyst and was thus considered unsuitable for public release. According to the Ohio EVEREST teams [Everest07], they were only permitted access to the Diebold/Premier private reports after their analysis was concluded, thus limiting its ability to help them in their work. Hart InterCivic simply forbade any access to the private reports on their system. This behavior on the part of the vendors is inexcusable. State-sponsored analysts operate under time and budget constraints and thus need access to the private work of their predecessors in order to more quickly get up to speed on how these systems work. Vendors should not be in a position where they can inhibit the work of professional analysts whose job is to examine their systems, nor should they have any power to censor these studies prior to their publication. The public has a right to detailed information about the strengths and weaknesses of their voting systems. Professional security analysts, working together with state sponsors, have demonstrated the ability to strike an appropriate balance between public disclosure of the *existence and severity* of vulnerabilities while relegating the sort of *supporting details* that could only aid an attacker to private appendices.

Recommendations

If Texas is going to continue purchasing and allowing equipment from the vendors who are currently certified in this state, then it is going to need to perform radically stronger oversight of these vendors' operations and future plans. **Internal vendor processes and procedures, ranging from their defect tracking to their blue-sky future system designs, need to be opened to state scrutiny and feedback.** This will be the only way to ensure that these vendors are seriously addressing the concerns that others and we have raised. If, for example, you were to demote current voting machines to a "provisional" status, pending vendor improvements, you should be able to have some confidence whether vendors are diligently fixing their systems or whether they will simply come back in two years and press for extensions. If a vendor is visibly failing to make progress, then counties using its equipment should be able to plan an orderly transition to other equipment.

Indeed, present-generation DRE systems have unacceptable security risks that cannot be mitigated simply through better election operations and procedures. California has taken the step of **limiting DREs to one per precinct**, to ensure accessible voting, while having most voters using paper ballots. That would be a prudent step to take here as well.

Electronic tabulation of paper ballots still has its security risks, but these can be mitigated with **hand audits of the paper ballots**, which can be conducted between the completion of the election and the certification of the final election results. Such audits involve randomly sampling ballots, by hand, and comparing them statistically to the electronic results. These audits can be made more accurate if the ballot tabulator were to stamp a serial number on the ballot (i.e., a number which the voter cannot see, but which is recorded both electronically and on paper). This would allow for one-to-one audits of electronic and paper records, greatly reducing the amount of effort necessary to conduct an audit.

Human factors research has shown significant variances across different voting technologies and different features of voting systems. **Human subject tests should become part of the state's certification process**, conducted by the state's board of election examiners with test subjects from the general population. Such tests would give local election officials more objective data to use when making purchasing decisions. These tests would also give the state concrete, measurable metrics on which vendors can be compared (or required to improve). Likewise, such tests would be able to determine best practices for how ballots should be designed and how other features of these systems should be configured.

Based on the current human factors literature, we can recommend the **elimination of straight ticket voting**. The straight ticket feature simply confuses voters, causing as many as 3% of ballots to have errors. Many other states, including California, forbid straight ticket voting features. For similar reasons, many other states **rotate the order of candidates on the ballot**, to avoid a statistically significant improvement to the votes garnered by the top candidate on the list. Texas should adopt these reforms.

Lastly, I want to give a word of hope for future-generation DRE systems, which could be designed using sophisticated cryptographic and other techniques to provide a level security and auditability not available with any voting system on the market today. Getting these techniques from the research world to the voting system industry won't happen automatically. Legislation or regulation can require DREs to have "end to end" verification properties, and provide a high bar for vendors to prove their systems meet these goals. With such systems, we are no longer required to trust that the "black box" operates correctly. Instead, we can challenge these systems, during the election, to *prove* that they are operating correctly. Research prototypes, such as our own open-source VoteBox system [SKW08], have these features and could form the basis for subsequent commercial systems achieving better security and auditability, both for traditional elections as well as remote and overseas voting [SW08]. The federal VVSG 2007/2008 standards have an "innovation class" that considers how such systems might be certified and tested, but none of this really matters until vendors bring products like this to the market. If the current vendors have no plans to produce better voting systems, then Texas should consider commissioning its own systems, from scratch.

Citations

[SKW08] Daniel R. Sandler, Kyle Derr, and Dan S. Wallach. VoteBox: a tamper-evident, verifiable electronic voting system. *Proceedings of the 17th USENIX Security Symposium (USENIX Security '08)* (San Jose, California), July 2008.
http://www.cs.rice.edu/~dsandler/pub/sandler08votebox.pdf
See also, http://votebox.cs.rice.edu

[SW08] Daniel R. Sandler and Dan S. Wallach. The case for networked remote voting precincts. *Proceedings of the 3rd USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '08)* (San Jose, California), July 2008.
http://www.cs.rice.edu/~dsandler/pub/sandler08remote-voting.pdf

[HMH+08] Paul S. Herrnson, Richard G. Niemi, Michael J. Hanmer, Benjamin B. Bederson, Frederick C. Conrad, and Michael W. Traugott. Voting Technology: The Not-So-Simple Act of Casting a Ballot. Brookings Institution Press (Washington, D.C.), 2008.

[EGB+08] Everett, S. P., Greene, K. K., Byrne, M. D., Wallach, D. S., Derr, K., Sandler, D., & Torous, T. Electronic voting machines versus traditional methods: Improved preference, similar performance. *Human Factors in Computing Systems: Proceedings of CHI 2008* (Florence, Italy), April 2008. http://chil.rice.edu/research/pdf/EverettGreeneBWDST_08.pdf

[F08] Edward Felten. NJ Election Day: Voting Machine Status. *Freedom to Tinker*, June 2008.
http://freedom-to-tinker.com/blog/felten/nj-election-day-voting-machine-status

[FHHL08] Laurin Frisina, Michael C. Herron, James Honaker, Jeffrey B. Lewis. Ballot Formats, Touchscreens, and Undervotes: A Study of the 2006 Midterm Elections in Florida. *Election Law Journal*. 7(1): 25-47, March 2008.
http://www.liebertonline.com/doi/abs/10.1089/elj.2008.7103

[AL08] Arlene Ash and John Lamperti. Florida 2006: Can Statistics Tell us Who Won Congressional District-13? *Chance*. 21(2): 18-26, Spring 2008.
http://www.amstat.org/PUBLICATIONS/chance/pdfs/199.featured.pdf

[Everest07] EVEREST Testing Reports, December 2007.
http://www.sos.state.oh.us/elections/voterinformation/equipment/VotingSystemReviewFindings/EVERESTtestingReports.aspx

[B07] Alan Bernstein. Election fixes stir worries on ballot security / Some fearful computer codes are vulnerable. *Houston Chronicle*, Page 1A, November 14, 2007.
http://www.chron.com/CDA/archives/archive.mpl?id=2007_4461049

[CA-Hart07] Srinivas Inguva, Eric Rescorla, Hovav Shacham, and Dan S. Wallach. Source Code Review of the Hart InterCivic Voting System. *Report commissioned as part of the California Secretary of State's Top-To-Bottom Review of California voting systems* (Sacramento, California), July 2007.
http://www.sos.ca.gov/elections/voting_systems/ttbr/Hart-source-public.pdf

[CA-Diebold07] Joseph A. Calandrino, Ariel J. Feldman, J. Alex Halderman, David Wagner, Harlan Yu, William P. Zeller. Source Code Review of the Diebold Voting System. *Report commissioned as part of the California Secretary of State's Top-To-Bottom Review of California voting systems* (Sacramento, California), July 2007.
http://www.sos.ca.gov/elections/voting_systems/ttbr/diebold-source-public-jul29.pdf

[E07] Everett, S. P. *The Usability of Electronic Voting Machines and How Votes Can Be Changed Without Detection.* Doctoral disseration, Rice University (Houston, Texas), May 2007.
http://chil.rice.edu/research/pdf/EverettDissertation.pdf

[BGE07] Byrne, M. D., Greene, K. K., & Everett, S. P. Usability of voting systems: Baseline data for paper, punch cards, and lever machines. *Human Factors in Computing Systems: Proceedings of CHI 2007* (San Jose, California), April 2007.
http://chil.rice.edu/research/pdf/ByrneGreeneE_07.pdf

[W06] Dan S. Wallach. Security and Reliability of Webb County's ES&S Voting System and the March '06 Primary Election. Expert report in *Flores v. Lopez* (Laredo, Texas), May 2006.
http://accurate-voting.org/wp-content/uploads/2006/09/webb-report2.pdf

[W08] Dan S. Wallach. Voting System Risk Assessment via Computational Complexity Analysis. *William and Mary Bill of Rights Journal*, Vol. 19, to appear December 2008.
http://accurate-voting.org/wp-content/uploads/2008/08/risk-eval-final.pdf

[KSRW04] Tadayoshi Kohno, Adam Stubblefield, Aviel D. Rubin, and Dan S. Wallach. Analysis of an Electronic Voting Machine. IEEE Symposium on Security and Privacy (Oakland, California), May 2004.
http://avirubin.com/vote/analysis/