VoteTrustUSA
National Coalition for Election Integrity

E-Voter Education Project

DRE Reliability: Failure by Design?

by Howard Stanislevic
Research Consultant
VoteTrustUSA E-Voter Education Project

March 13, 2006

www.VoteTrustUSA.org

DRE Reliability: Failure by Design?

by Howard Stanislevic
Research Consultant
VoteTrustUSA E-Voter Education Project

This is the first in a series of presentations which will explain the risks and benefits of electronic voting (e-voting) systems to candidates, elected officials and the public at large.

Abstract – *The Election Assistance Commission's Voluntary Voting Systems Guidelines (VVSG) are developed by a Technical Guidelines Development Committee[1] that includes participation by the International Institute of Electrical and Electronics Engineers (IEEE) – a 375,000-member non-profit professional organization whose code of ethics states that its members shall "accept responsibility in making engineering decisions consistent with the safety, health and welfare of the public." We show herein that the VVSG Reliability standard for e-voting systems, expressed as a Mean Time Between Failures (MTBF) of only 163 hours, is woefully inadequate, resulting in voting systems of lower reliability than mechanical lever voting machines in use for over 40 years. This standard allows 9.2% of all e-voting systems to fail in any 15-hour Election Day, and a much higher failure rate during the extended "Early Voting" periods now being implemented in many states.*

At a Voting Systems Testing Summit held by California Secretary of State Bruce McPherson last November in Sacramento, Carnegie Mellon University computer expert Michael Shamos, a voting-systems certification official for the state of Pennsylvania, is reported to have asked publicly, "Why are we not in an uproar about the failure of (touch-screen voting) systems?…I have good reason to believe that 10 percent of systems are failing on Election Day. That's an unbelievable number."[2]

An examination of the Reliability standard approved by the federal Election Assistance Commission (EAC) shows that this spec is highly consistent with the 10% failure rate Dr. Shamos finds so outrageous. As to why there is no uproar, we can only assume it's because the American electorate has not been properly informed concerning this issue – until now.

Failure *Is* An Option

It is common knowledge among engineers that the reliability of electronic devices and systems is stated as the mean (average) time between failures or MTBF. Today's technology is so reliable that MTBFs of 1,000,000 hours (114 YEARS) are not uncommon for some products. Here are typical MTBFs for some common devices and systems:

[1] http://www.eac.gov/tgdc.asp
[2] Ian Hoffman, Uncertainty Clouds Future of E-vote Tests, Oakland Tribune, December 01, 2005, http://www.votersunite.org/article.asp?id=6414

Product	MTBF: Hours	Years
Computer Hard Drive	1,000,000	114
Thin Client PC (no disk drives)	170,000	19
Traffic Light (LED)	100,000	11
VCR (Security System)	70,000	8
TV/Monitor	45,000	5
DVD Player	40,000	4.5
Standard PC	30,000	3.4
Touch Screen Display	30,000	3.4
Light bulb (compact fluorescent)	10,000	1.1
Light bulb (incandescent)	1,000	0.11

But how are products actually tested for such long periods of time? And with today's fast pace of product obsolescence, why would anyone care if they actually lasted more than a few years? What does MTBF really mean and how is it used to certify the reliability of e-voting systems?

The answer is relatively simple, but it should also be of grave concern to all Americans: MTBF numbers are based on testing lots of devices for short periods of time, and in the case of e-voting systems, most of this testing is actually performed by voters on Election Day.

The EAC has recently approved the latest (2005) version of the Voluntary Voting System Guidelines (VVSG), scheduled to take effect in 2007. But these standards contain the same Reliability spec for Direct Recording Electronic voting machines (commonly known as DREs) as they did back in 1990. This wouldn't be so bad if this standard actually required highly reliable voting systems (at least as good as the ones they're replacing), but for the last 15 years, the Reliability standard that DREs have had to meet, first set by the FEC and under the Help America Vote Act (HAVA) by the EAC, is a mean time between failures of **only 163 hours.**

And here is the current (2002) spec to prove it:

Volume 1, Performance Standards
Sect. 3 - Hardware
3.4.3: Reliability
The reliability of voting system devices shall be measured as [M]ean [T]ime [B]etween Failure[s] (MTBF) for the system submitted for testing. MBTF [*sic*] is defined as the value of the ratio of operating time to the number of failures which have occurred in the specified time interval. A typical system operations scenario consist[s] of approximately 45 hours of equipment operation, consisting of 30 hours of equipment set-up and readiness testing and 15 hours of elections operations. For the purpose of demonstrating

compliance with this requirement, a failure is defined as any event which results in either the:

a. Loss of one or more functions; or
b. Degradation of performance such that the device is unable to perform its intended function for longer than 10 seconds.

The MTBF demonstrated during qualification testing shall be at least 163 hours.

It's important to note that by definition, any MTBF spec is only an average. While the above standard is repeated verbatim (except for some of the typos) in Sect. 4.3.3 of the latest (2005) version, the actual National Certification Test Design Criteria mandated by the EAC in Volume II, Appendix C, Sect. C.4 of the 2002 VVSG actually allows 4 failures after 409 hours of testing (MTBF of 102.5 hours), while the same section of the 2005 VVSG allows 6 failures after 466 hours (MTBF of 77.8 hours).

To appreciate just how lax these standards are, one only needs to review the above table in the present document. Even an old fashioned incandescent light bulb may fail less often than a DRE, and standard PCs, touch screen displays and hard drives are expected to fail hundreds of times less often.

Note that the VVSG standard makes no distinction between failures which voters would be aware of (e.g., screen freezes) and those that voters would not notice at all (e.g., failure to record votes in memory). Nor is there any differentiation made between machines that can be repaired in a timely manner on Election Day and those requiring replacement.

Although this spec also applies to paper ballot/optical scan (mark-sense) systems, please consider that while ballot scanners are as important to the vote counting process as DREs, the scanners have nothing to do with vote casting. If a scanner fails at a polling place, a voter has a choice of either leaving her paper ballot safely in the custody of elections officials to be counted later (usually the same day), or waiting until the scanner is repaired or replaced. Either way, the vote can be recorded and cast; no voter is forced to wait longer or be disenfranchised due to a scanner failure. However, such is not the case with a failed DRE which voters must rely on to record, cast and count their votes and even to create a voter-verified paper audit trail (VVPAT) if required by law.

Consider now how this standard could affect an actual election:
The MTBF of 163 hours can be realized by running a number of DREs for a short amount of time (such as a 15-hour election day), then multiplying the time by the number of DREs to get the total machine-hours, and dividing that by the total number of failures:

(1) $$M = H \times N / F$$

where *M* is the MTBF in hours, *H* is the number of hours the polls are open, *N* is the number of DREs in a jurisdiction or state, and *F* is the number of DREs that failed during the election.

We can calculate the number of DREs that are permitted to fail on a given Election Day for a given MTBF with the following equations, which solve for *F* from equation (1):

$$M \times F = H \times N$$

(2) $$F = H \times N / M$$

In other words, the number of failed DREs allowed *(F)* is equal to the number of hours in the Election Day multiplied by the number of DREs in the jurisdiction, divided by the MTBF of 163 hours. With this MTBF, the failure rate allowed during any 15-hour period *(F / N x 100)* will always equal **9.2%** or **one out of every 11 DREs.**

Let's use the city of New York as an example:

Big Problems in the Big Apple

New York City deployed 7,300 lever machines in the 2004 general election. The polls were open for 15 hours. Let's use the federal standard to calculate the number of DREs that would have been permitted to fail on Election Day using equation (2). Assuming a one-for-one replacement of lever machines with DREs:

$$F = 15 \times 7{,}300 / 163 = 672$$

Six hundred and seventy two voting machine failures? Why, that's 9.2% of the 7,300 machines in NYC. In fact, 9.2% is the failure rate allowed in any 15-hour period for virtually any number of machines under the federal Reliability standard.

Now according to the NYC Board of Elections, only 469 lever machines (6.4%) actually failed during the 2004 general election. Thus, the failure rate allowed for DREs by the 21st-century VVSG Reliability spec exceeds the actual failure rate of 40-year-old lever machines by 44%. Furthermore, the residual vote rate for the lever machines was about 0.9% in 2004[3] -- less than the national average for DREs.[4]

[3] http://vote.nyc.ny.us/pdf/results/2004/general/g2004recaps.pdf
[4] David C. Kimball, PhD, University of Missouri-St. Louis, Summary Tables on Voting Technology and Residual Vote Rates, Dec. 14, 2005, http://www.umsl.edu/~kimballd/rtables.pdf

Only 20 lever machines (0.27%) had to be replaced in the 2004 general election in NYC.[5] The vast majority of the failed machines (96% of them) were repaired after about 2½ hours on Election Day. And in 2005, there were only 7 replacements required out of 7,347 lever machines – less than 0.1%.[6]

To investigate further, we calculated the MTBF for the 40-year-old NYC lever machines based on their total failure rates (repairs + replacement) and replacements only, using equation *(1)*:

> ***2004 Total Failures:*** *M = 7,300 x 15 / 469 = MTBF of 233.5 Hours*
> ***2004 Replacements:*** *M = 7,300 x 15 / 20 = MTBF of 5,475 Hours*
> ***2005 Replacements:*** *M = 7,347 x 15 / 7 = MTBF of 15,743.6 Hours*

Amazingly, the federal guidelines would have allowed 44% more total DRE failures and 33 times more DRE replacements than the actual lever machine failures and replacements in the 2004 Presidential election -- or a DRE failure or replacement every 80 seconds in the city of New York.

We also know that lever machines are not vulnerable to problems such as faulty smart card readers, software bugs, touch screen miscalibration, vote switching, or electrostatic discharge. Indeed, a voter or poll worker walking across a carpet can generate electrostatic voltages far in excess of those used to test DREs according to the VVSG,[7] while nothing short of a direct lightning strike would affect a lever machine.

According to a New York attorney with whom we have consulted, New York State must meet not only the low standards set by the EAC under the federal Help America Vote Act, but the higher standard set by the state's Constitution. The equal protection clauses in Article I (The Bill of Rights), § 1 and § 11 require not only that any replacement voting system be equal for each citizen of the state, but that it also provide safeguards equal to the system it replaces. A Reliability standard for DREs comparable to the reliability of lever machines would therefore require a much higher MTBF than the 163 hours allowed by the federal standards.

Of course, the Reliability issue is not unique to NY. Here's a sample of "DRE states" with their allowable quantities and frequencies of machine failures based on the VVSG's 163-hour MTBF and a 15-hour Election Day:

---

[5] John Ravitz, Executive Dir. NYC BoE, WNYC Radio, The Brian Lehrer Show, "RIP Lever Machine", Nov. 8, 2005, http://audio.wnyc.org/bl/bl110805d.mp3
[6] Ravitz, Testimony before the City Council Government Operations Committee, Nov. 21, 2005
[7] http://www.esda.org/basics/part1.cfm

State	Approx. # of DREs	# of Failures = 1 DRE Failure Every…	
GA	26,000	2,393	23 seconds
MD	16,000	1,472	37 seconds
NJ	7,500	690	78 seconds
NC	7,400	681	79 seconds
NM	2,500	230	234 seconds

Since the federal standard does not differentiate between failures requiring repairs and those requiring machine replacements, it's instructive to consider some actual Election Day 2004 experience to see how the DREs stacked up against the levers.

When the Right to Vote Goes Wrong

Even though the EAC standard seems rather lax, it could be argued that vendors may exceed this spec and provide the American electorate with a highly reliable voting system. However, experience in the state of Maryland in the 2004 general election has shown that the actual failure rate of DRE systems may have exceeded even that predicted by the weak federal standard.[8]

TrueVoteMD had observers in 108 of the state's 1,787 precincts. According to Stanley A. Klein, DSci, MSEE – an electrical engineer and member of the TrueVoteMD Technical Committee, as well as the National Committee for Voting Integrity, and the IEEE's Project 1583 Committee tasked with the development of standards for voting equipment[9] -- out of approximately 967 DREs, there were 111 failures on Election Day 2004 that do not even include failures of smart card encoders, incomplete ballot definitions, or any failures that may have been unobservable to the voters. This turns out to be 11.4% of the machines in these precincts.[10]

Klein explained that the current flawed standard just flowed down from the 1990 FEC guidelines without any attention being given to it until he raised the issue in his comments on the 2002 standard. And there were other dissenting IEEE members as well. Klein believes that the decision to retain the lax standard was driven by the goal of minimizing the cost of DRE testing borne by the vendors. He told EVEP, "Proper reliability testing takes much more time and effort."

The members of TrueVoteMD weren't the only ones in the state who observed the high DRE failure rates in 2004; the Montgomery County Board of Elections publicly cited them in their report, "2004 Presidential General Election Review - Lessons Learned."[11] Out of 2,597 machines deployed, there were a total of 189

[8] When the Right to Vote Goes Wrong: Maryland Citizens Tell the Story of Election Day 2004, http://truevotemd.org/Election_Report.pdf
[9] http://standards.ieee.org/announcements/votingcall.html
[10] http://www.vote.nist.gov/ecposstatements/comment-memo-5_3_2.pdf
[11] http://www.truevotemd.org/Resources/Lessons_Learned.pdf

confirmed failures or 7.3%. Unlike the NYC lever machines however, nearly all these DRE failures required machine replacement.

The board had a total of 95 spare DREs, but because 5 of those had also failed, the available spares were reduced to 90. On the morning of the election, 58 DREs failed to boot up, showing a Ballot Exception Error. These were unusable and were immediately taken out of service. Another 106 DREs exhibited screen freezes. In the Board of Elections staff's opinion this was the most serious of errors, leaving the voters and election judges unable to verify that votes were in fact being counted. Therefore at the very least, 164 DREs (6.3%) should have been replaced, but there were only 90 available spares. An additional 25 failures (1%) occurred due to card readers, printers and power problems, and 122 additional machines (4.7%) may have captured fewer votes compared to other machines in the same polling places, making them suspect.

According to the VVSG standards, the failure rate is derived from a random exponential probability distribution. This means that no one can predict when or where the failures will occur. So even enough spares to replace 9.2% of deployed machines and the 9.2% of the spares that are also allowed to fail (about 10% total replacements) may not be sufficient because no one can know in advance where to position the spares. This has led some to believe that vendors are happy to have a poor Reliability standard so elections officials will be forced to buy more spare equipment. Indeed, in the above report, the Montgomery County BoE suggested at least one spare DRE per polling place if their equipment problems could not be resolved.

Once again, by contrast, in the same election, the NYC Board of Elections had to replace only 20 of its 7,300 lever machines (0.27%), so clearly the old clunkers were much more reliable than Montgomery County, MD's state of the art DREs.

Untimely Voting

So far, we have only examined the failure rates allowed for DREs on Election Day, but increasingly in many states, DREs are being deployed for early or absentee voting that lasts for days or even weeks before an election. In the state of Georgia for example, most of its 159 counties have at least one DRE in operation during normal business hours for five days prior to Election Day. That's at least 40 hours per DRE. An MTBF of 163 hours predicts a failure rate of almost 25% during this time period, instead of the already unacceptable 9.2% in a 15-hour Election Day.

99% Uptime?

We've seen that it's not unreasonable to expect 10% of DREs to fail on every Election Day, so are there any requirements in the standards to get them fixed?

The answer is: Yes – sort of. But this requirement, known as Availability, is also based on the lax Reliability (MTBF) spec. Here's how it works:

DRE Availability has to be at least 99% under Sect. 3.4.5 and Sect. 4.3.5 of the 2002 and 2005 standards respectively, which on the surface appears to be somewhat better than the 90.8% Reliability (9.2% failure rate) previously cited. But here's the catch: The definition of Availability simply adds the Mean Time to Repair the machines (MTTR) to the MTBF number. Since the MTBF is 163 hours, to achieve "99% Availability" according to this standard, the 9.2% of the machines permitted to be broken on Election Day must be fixed within 1.63 hours. Here's the formula for this standard:

(3) $$Availability = MTBF / (MTBF + MTTR)$$

A 1.63-hour MTTR (average repair time of 1 hour and 38 minutes) gives us the allowed 99% Availability:

$$Availability = 163 / 164.63 = .99$$

However, the effect of this spec is simply to allow an average downtime of an hour and 38 minutes per failed voting machine during which voters cannot cast their votes. Readers should therefore be aware that "99% Availability" does not mean that failed voting machines must somehow be available 99% of the time on Election Day. The standard allows 1.63 hours for the machines to be repaired, which is 10.87% of a 15-hour Election Day or 13.58% of the shorter 12-hour Election Days found in many states. This assumes it's even legal for a technician to gain access to a broken voting machine during an election. In some states, this might be prohibited. It also assumes there are enough technicians available statewide to reach the failed units in time and that the units can be repaired or replaced in a timely manner.

In round numbers then, the federal standards permit almost 10% of machines to be down at least 10% of the time every Election Day, with a much higher failure rate (e.g., 25%) during the longer early voting periods.

Another interesting fact is that as MTBF increases, longer repair times and/or higher availability can be used to satisfy equation (3). For example, a longer MTBF of 15,000 hours (proposed by Stanley Klein) – which is still less than that of a standard PC[12] -- with the same 1.63-hour MTTR, would result in an Availability of 99.99%. Now that's a spec worth aiming for.

[12] J. B. Miles, "Thin Clients", Government Computer News, October 2, 2000; Vol. 19 No. 29, http://www.gcn.com/vol19_no29/guide/3040-1.html

So what do the vendors say?

In industries other than e-voting, MTBF is often predicted by using either of two major standards: MIL-HDBK-217 (a US Department of Defense standard for reliability prediction) or Telcordia SR-332 (developed for commercial products by Bell Communications Research). In general, both of these time-tested standards add up the failure rates of all the components of a device to arrive at a probability that the device as a whole will fail. For example, if a device consists of 10 components, and each one has a 1% chance of failing during a given time interval, the device as a whole will have a 10% chance of failing during the same period. (MTBF can easily be calculated once a failure probability is known.)

But according to findings in a 2003 report issued by Ohio Secretary of State J. Kenneth Blackwell[13], nearly all of the e-voting vendors questioned seemed to have little or no understanding of the science of reliability engineering.

According to one vendor's response: "No vendor bidding the current generation of equipment can make iron-clad guarantees to life expectancy beyond full compliance with FEC standards."[14] But of course, that's only 163 hours.

Another vendor claiming an operational life span in excess of 10-15 years says they are "not certain what 'documentation' can be provided to support these projections."[15]

Instead of taking the failure rates of all components into account, yet another vendor simply used the component with the highest rate of failure, significantly overstating the MTBF of their DREs.[16]

Only one vendor actually cited a commonly accepted Reliability standard (MIL-HDBK-217) as a basis for their reliability prediction, but the details of their calculations were kept strictly confidential.[17]

It should of course be noted that no claim made by any vendor is a substitute for truly independent public testing and a high Reliability standard, neither of which exist in the federal certification process. This is one reason why the state of California has conducted it's own volume testing.

---

[13] Vendor Proposal Evaluation Findings Report & Addendum, Statewide Voting System(s), Ohio Secretary of State J. Kenneth Blackwell, Prepared by Nola Haug, Report Date: August 15, 2003, Addendum Date: September 10, 2003, http://www.sos.state.oh.us/sos/hava/findings091003.pdf
[14] Ibid, Pg. 140 ES&S
[15] Ibid, Pg. 157 Sequoia
[16] Ibid, Pg. 136 Diebold
[17] Ibid, Pg. 146 Hart/Intercivic

Turning Up the Volume

In a 5-hour test of 96 DREs, a MTBF of only 15 hours was observed, prompting the authors of this landmark report to state, "The failure to detect this fact during the [federal] ITA's testing process appears to be due to serious defects in the testing methodology specified by federal standards. One lesson of this analysis is that the testing performed during the federal qualification process is apparently inadequate to ensure that voting machines will be reliable enough for use in elections."[18]

It would seem that the federal certification process itself may be unreliable and in urgent need of repair. So when computer scientists, engineers, elections officials, candidates, legislators or voters express surprise and concern that their e-voting systems fail in high numbers on Election Day, they would be well advised to consult the Reliability standard such machines are expected to meet. Widespread voting system failure, if not by design, is certainly an option under this standard -- and should come as no surprise to anyone paying attention.

Failure Mitigation Strategies

The present work would not be complete without turning our attention to the task of failure mitigation. Consider now the challenges posed to the well-intentioned elections administrator:

The current and previous generations of DREs have been designed with the possibility of a 9.2% average failure rate in a 15-hour election day. In two years, the latest version of the VVSG, which allows even more failures during the testing process, will take effect. Such failures are unpredictable; they could occur randomly anywhere at any time; they may be obvious to the voters or go completely undetected -- there is simply no way to know in advance. Failures of less than 10 seconds in duration can occur in unlimited numbers and all of the above also applies to the pool of spare equipment purchased to mitigate these risks.

While VVPATs could provide some redundancy in the system to guard against lost votes, they are not always legally the ballots of record, nor are they always counted or even randomly audited. Harried or less educated voters may not even verify them. Elections officials could change this by law, regulation and voter education, but this is only a partial remedy because the VVPATs themselves are

[18] Analysis of Volume Testing of the AccuVote TSx / AccuView, Matt Bishop, Loretta Guarino, David Jefferson, David Wagner, Voting Systems Technology Assessment Advisory Board with assistance from statistician Michael Orkin (Managing Scientist, Exponent), October 11, 2005, http://www.ss.ca.gov/elections/voting_systems/vstaab_volume_test_report.pdf

produced by the DREs and therefore subject to the same poor reliability standard.

Mechanical lever machines are still an option for jurisdictions that haven't discarded them, but of course they do not meet the HAVA Accessibility requirement of one system per polling place to accommodate disabled voters.

As previously noted herein, paper ballot systems offer an acceptable solution since even a failed ballot scanner will not disenfranchise voters. Thanks to medical science, the predicted MTBF of a voter is currently over 70 years, and it's the voters – not the DREs -- that mark the paper ballots. Paper ballots are therefore inherently voter-verified and can be scanned, counted by hand, or both, allowing fully independent auditing. Clearly this mitigates the risk of equipment failure.

Meanwhile, the federal voting systems standards development process must be subject to investigation and oversight. The fact that a Reliability standard such as this has been allowed to remain on the books for over 15 years is cause for grave concern.