

Observations and Recommendations on Pre-election testing in Miami-Dade County

Douglas W. Jones
Department of Computer Science
University of Iowa
Iowa City, IA 52242
(jones@cs.uiowa.edu)

Sept 9, 2004

Summary

Miami-Dade County performed pre-election testing for their August 31st 2004 primary election on August 13th. Prior to these tests, draft procedures, dated 8/11/2004 were distributed by the Elections department, revised from preliminary documents distributed on 8/9/04. My understanding is that Orlando Suarez and Cathy Jackson were both involved in planning for these pre-election tests.

Overall, my impression of the testing was that it was, stronger than the testing undertaken by most jurisdictions and stronger than that required by the state of Florida, and yet, at the same time, neither as strong as was intended nor as strong as it could have been. The testing uncovered some real problems that ES&S, the manufacturer of the iVotronic should correct, but these problems do not prevent the use of the iVotronic in upcoming elections or call into doubt the results of elections cast on this machine.

My observations and recommendations are divided as follows:

1) Test Script Design and Unscripted Tests	1
2) Testing Overvoted Ballots	3
3) Sample Size versus Depth of Testing	4
4) Test Ballot Entry Procedures	4
5) Test Organization	5
6) The Event Log Bug	10
7) Security	11
8) Absentee Ballot Counting	15
9) Absentee Ballot Instructions	16
10) Openness and Observability	18
11) Touch Screen Calibration	20
12) Recommendations to ES&S for Improvements to the iVotronic and Unity	23

Some of these recommendations will clearly take time to implement, particularly those in the final section, addressed to Election Systems and Software. Others should be fairly easy to implement in time for the November election and the mid-October pre-election tests.

1) Test Script Design and Unscripted Tests

I have always held that test scripts where all of the candidates get the same number of votes are weak. Florida law on pre-election testing, Title IX, Chapter 101.5612, does not dictate a test pattern, other than requiring that the ability of the system to reject overvotes be tested. Rather, it focuses on the use of whatever test pattern is used -- verifying that the output from the voting system matches the inputs.

When, on the morning of the test, Paul Kraft recommended against mingling unscripted test votes with scripted test votes, I pointed out that this would result in a loss of test coverage. I do not want to argue "what should have been done" at the August 13 test, but rather, how to move back to better test coverage without creating the problems that led Paul Kraft to object. I can think of three paths to this goal:

1) Conduct unscripted voting on top of scripted voting, as was originally proposed on August 13, but add a step in which the test ballots created by the public would be examined by members of the audit team to update their predicted outcomes for the election to include these unscripted ballots. This requires very careful manual record keeping, but permits the full extent of public involvement that was originally proposed.

2) Invite members of the public to create unscripted test ballots and then add them to the script. At the appointed time for public participation, voters would be given test ballots and asked to mark them, and then, after the votes in these ballots were, manually, added to the expected test outcome, these ballots would be simply added in to the ballots used for the scripted testing. Thus, the public would no longer be invited to directly cast unscripted votes during the tests.

3) The designers of the script, without involving the public at all, could throw coins or otherwise select at random to add extra ballots to the basic script. Again, these randomly voted ballots should be included in the test plan so that, once the script is voted, the results can be reconciled against the expected results.

From a statistical viewpoint these different options are comparable. The only real difference lies in the extent of public involvement. While public involvement in the testing may build significant public confidence, it is not really necessary.

2) Testing Overvoted Ballots

Florida law requires that both the DRE and absentee systems be tested for their ability to handle overvotes. In the case of the DRE machines, the test assures that the machine contains appropriate interlocks to prevent the casting of more than the allowed votes in a race. In the case of the absentee system, the test assures that overvotes are not counted.

In the August 13 test, Miami made a good faith effort to follow these rules by including a fully blank ballot and a fully overvoted ballot in the test script for each ballot style. The results, though, were troublesome. The very same test script was used for the absentee system as for the DRE system, but the totals were not the same, because attempts to enter overvotes into the DRE machines were not handled in the same way that attempts to enter overvotes into the absentee system were handled. There is nothing wrong with this, except that it led to the awkward need to include, in the official report of the test, the disclaimer that the number of ballots counted on the DRE and absentee systems were different because of the different way in which overvotes were handled.

The following proposal for testing overvoted ballots should improve on this:

1) In the test script, prepared on optical mark-sense ballots, as was done on August 13, where ballots are overvoted, some of the overvotes should be in the form of corrections, that is, filled in ovals with an X through them, as indicated in the ballot instructions for making corrections.

2) When voting an overvoted race on the DRE system, if the script ballot has overvotes that include a correction, attempt to overvote, and then clear out all selections with the exception of

the selection(s) that are not marked with an X. As a result, this overvoted ballot, after attempting to overvote, will be counted as a legitimate vote. If no alternatives are X'd out in an overvoted race, after attempting to overvote, all selections in that race should be cleared and no candidate should be selected.

3) In counting absentee ballots, the tabulating system should be set to sort out overvoted ballots, and those with corrections should then be handled by the canvassing board or their designees as they would on election day, creating true copies embodying the intent of the overvoted ballots for tabulation.

This procedure should lead to identical totals on both the DRE and absentee systems, and it is a more realistic test of the full election day procedures. If done using our test scripts from August 13, this procedure would simply involve casting a blank ballot for every fully overvoted test ballot, but the above rules also allows the possibility of distributing overvoted race tests over other test ballots instead of lumping them unrealistically on single grossly overvoted test ballots.

3) Sample Size versus Depth of Testing

The August 13 test in Miami-Dade County took too long. Some of this can be attributed to the organization of the test and to unexpected problems that caused unavoidable delays, but some of this might be the result of testing more machines than necessary. It is fair to speculate that testing a smaller number of machines more intensively might have been more valuable than testing a large number of machines with shallower test coverage.

To the extent that the sample size used in the August 13 test was a direct result of the attempt to gain complete and representative coverage of the mix of ballot styles being used, then I would not question the sample size. I believe that full coverage of the marriage between ballot styles and voting machines is important enough to merit going beyond the state mandated 2 percent.

If, on the other hand, additional machines were included in the test for the sole reason of increasing the sample size beyond that required by law, I would question that decision and suggest that the same effort could be more effectively used to increase the intensity of testing on a smaller sample of the machines.

4) Test Ballot Entry Procedures

The procedure used for entering test ballots on the DRE machines worked fairly well. Each test ballot was entered by a voter, under the supervision of a scribe, where the scribe was responsible for reading the ballot verification screens and checking that the ballot, as voted, matched the script. With two people doing this, errors in entering the test ballots were extremely rare. Unfortunately, extremely rare does not mean nonexistent.

In any effort of this scale, it is reasonable to make allowances for human error. I recall one presentation by a DRE system vendor in the mid 1990s where the vendor's sales rep said: "You cannot expect to do a realistic test of a DRE system by entering votes by hand. It is so hard to do this that if you enter a realistic number of ballots, you will make so many mistakes that your test is useless. Therefore, you have to rely on the testing done by the ITAs and on the self test scripts built into the machines."

Clearly, this assessment of DRE voting system testing overstates the difficulty. The number of errors made in entering test ballots using the procedures used on August 13 was small enough that we can assert, with confidence, that the test procedure with voter and scribe is a sound procedure.

In looking at the errors that were made, it is clear that many could have been avoided by careful test script design! I did not find a complete list of errors, and this raises concerns in my mind about the quality of record keeping for these tests. In listening to discussion of the errors, some categories were easy to recognize:

1) Failure to vote the entirely blank test ballot included in the script. In this case, both the scribe and test voter had to agree to do this, despite their instructions. This failure only changes the number of ballots counted, not the totals for any candidates, so it is natural for a test team to make this mistake.

The frequency of this mistake might have been avoided by any of several strategies: Consider, for example, writing on the blank ballot "vote one entirely blank ballot". The writing would have to be using markings that are not sensed by the mark-sense scanner, consider using highlighter pens to do this, or consider writing down the center of the column using a bold red marker. Or, use a post-it note with the special instructions, or attach the special instructions with a paperclip, or some combination of these.

I believe that the inclusion of entirely blank ballots in the test is not inappropriate, but that it is just as reasonable to scatter undervoted races throughout the ballot. The test script design used on August 13 did this for most races, because the number of test ballots was driven by those races that had the largest number of candidates. Those large races, however, never had any undervotes except on the entirely blank ballots. By transferring one test vote from one of the other test ballots to a totally blank ballot, the same test coverage would be achieved without any totally blank ballots, thus reducing the number of special cases dealt with by the testers.

2) Entry of the fully overvoted ballot as a blank ballot. In fact, this is the procedure recommended above, in the discussion of testing overvoted ballots, but in the August 13 tests, some test teams did not attempt to vote their totally overvoted test ballots, while others cast blank ballots for this.

This error would have been far less likely if special instructions for testing overvoted ballots were marked on the ballot, as suggested above for testing blank ballots. The special procedures for overvote testing suggested above would also clarify the process.

3) Voting on the wrong ballot style. Some test ballots contained only undervotes for the partisan races, with the only non-blank issues on the ballot being in nonpartisan races. As a result, voting on the wrong ballot style had no effect on the vote totals of any race, but only on the number of each style of ballot counted. It is easy to understand how a test team in a hurry could overlook this error.

I believe that the ES&S iVotronic is predisposed to this error because the touch-screen targets for ballot style selection are very small and closely spaced compared to the touch-screen targets for voting. Selecting the wrong style is, therefore, easy, and unlike voting, once you select the wrong style, the only way to back up is to cancel the selected ballot and start over or to find a ballot in the test script that matches the style that was accidentally selected.

During testing, roving supervisors reminded test teams that, during both voting and ballot review, blank races should be read off and declared to be blank instead of merely skipping down to the next race containing a vote. Once scolded, it is unlikely that a test team would overlook having selected the wrong ballot style, so the key to avoiding this problem in testing is probably to ensure that testing teams are properly coached to read, out loud, the word *undervote* for each race in which no vote is cast.

Even with these measures, there will probably be some mistakes. Because testing errors can bring into question the entire test, it is worth considering alternative procedures for reducing the frequency of testing errors, or if not reducing their frequency, producing objective evidence of them. Here are some ideas:

1) 3-person teams. Replace the voter-scribe teams with voter-scribe-checker teams. The current procedure in Miami-Dade County is to have the scribe read the scripted ballot to the voter, who votes it, and then the scribe and voter change places so the scribe can read the verification screen to the voter, who checks it against the scripted ballot. Only after this procedure has been conducted successfully does the voter cast the ballot.

One way to reduce the frequency of error is to add another person to this team, call them the checker. Initially, the scribe would read the ballot to the voter while the checker monitors, and then the checker would read the verification screens back to the voter for checking, while the scribe monitors. If all three agree, the voter would cast the ballot.

This scheme has the benefit of adding one more level of oversight to each test ballot cast, but it poses two problems: First, it increases the cost of casting each test ballot by 50%. Second, it increases the number of people attempting to all look at the touch screen at the same time. The privacy booths are designed to prevent two people from looking at once, and adding extra people in the aisle space between machines under test would add another dimension to the confusion of the entire process.

2) Use of a camera to record the testing. Paul Kraft suggested this on the morning of the 13th, and it has been frequently suggested by others for DRE testing. Use of a camera to record the entry of test votes is very easy, but use of the recording is not! If conventional video recordings are used, reviewing the record to search for the cause of a discrepancy it likely to take as long as it took to enter the original test votes.

If one camera is allotted to every voting machine under test, it is easy to find the record of a particular test, but the cost of camera management is maximized. If, on the other hand, one camera is assigned to each testing team, far fewer cameras are involved, but there is the new problem of moving cameras from machine to machine, and the new problem of indexing all the different tests seen by one camera so that they can be recalled if there is a question.

There is also the question of whether it is better to use a video camera running continuously or to use the camera to record shots of each verification screen. The former leads to far more work if there is ever a need to review the record of the test, while the latter imposes a burden on the testing team, requiring them to snap a shot of each verification screen, and providing the opportunity to forget to do so, adding further complexity to the testing process.

3) A change in methodology. Instead of voting the test script from pre-marked ballots, present the test script in a compact form and have the testing team transcribe the votes from the verification screen to paper ballots. With this scheme, the paper ballots become a manually created record of the test ballots that were entered into the machine, available to be compared with the script and to be compared with the output from the machine.

This model could be improved by using two teams, one to enter and verify the votes from the script, and one to transcribe from the verification screen to paper ballots before casting the vote. The first team would not be allowed to cast the ballot, while the second team would not be allowed to change any votes and would be the ones to cast the ballot.

Because these teams are independent, we have three records of the vote: What the script said it should have been, what the machine recorded, and what the recording team transcribed to

paper. If all three agree, all is well. If the machine's record differs from the other two, the machine is likely to be in error. If either the script or the transcribed record match the machine, but the other paper record is different, it is most probable that one of the teams made a mistake.

I find this third model more appealing than using video cameras, but organizing a test using this model will require a different organization of the testing process than the one used on August 13. I will discuss this in the next section. It is also worth noting that this approach to recording the test can be used in conjunction with cameras, if desired, and that this approach to recording the test is applicable to unscripted parallel testing as well as being applicable to pre-election testing.

5) Test Organization

The fact that the testing on August 13 lasted until long after midnight troubles me for several reasons. Foremost among these is that these tests are, under Florida law, public tests, and therefore, public observability is essential. Public observation is extremely difficult when the test runs onward until such late hours; it is encouraging that, despite this, one member of the public, Dan McCrea of the Miami-Dade Election Reform Coalition, was sufficiently interested to stay to the very end! In addition, of course, there are questions of overtime pay, the increased likelihood of human error and the sheer inconvenience of running tests that take so long.

During the tests, there were many points when it seemed that large numbers of people were sitting around waiting for a few people to do something. This also leads me to think about alternate ways of conducting the tests that would both speed the process and make better use of manpower.

At times, the process looked chaotic, and it is clear that attempts to accelerate the testing could make it appear even more chaotic. As the August 13 tests wound down, though, a sense of order emerged from the chaos, and at one point, Dan McCrea made a comment that sheds some light on this, noting that when you see order emerge out of chaos, there is a good chance that, under the appearance of chaos, there was something orderly going on all the time.

As organized, the testing on August 13 was run with little or no overlap between different elements of the test. So, all machines involved in the testing were powered up and all zero tapes were printed before any test ballots were cast. Then, and only then, test ballots were cast on the first group of machines (those in the loading dock area), and only when these machines were completed did testing move to the second batch of machines, in the aisle between the storage racks. Only after all test ballots were entered were the polls closed on the machines under test, and only when all poll closing reports had been printed was the process of transporting PEBs to the collection center begun.

I speculate, although it was never explicitly explained to me, that this procedure was followed in order to keep all activity localized, so that the canvassing board could directly oversee each step in the testing. Had the tests been completed in one 8-hour shift, this model would make some sense, but with testing extending until 2:30 AM, 2 of the 3 members of the canvassing board left before the tests were finished. As a result, the final part of the testing was not as closely overseen by the canvassing board as might have been hoped for.

Here is a proposal for an alternative testing model that I hope would allow multiple phases of the test to be carried out at the same time. I do not want to suggest that this is the only way to accelerate the testing, but I hope that some of these ideas are useful. I also hope that, whatever procedure is used for testing in the future is documented, in advance, at a level of detail greater than I have given here:

Pre-test: Each machine, in its booth, is matched with a manilla folder. Attached to the outside of the manilla folder is a check-list of testing steps, where the time of day of each testing step is recorded. Inside the folder is the test script. In addition, each machine has an empty envelope large enough to hold the folder and serve as the repository for the permanent record of the test, including records of voted test ballots, poll opening tapes, poll closing tapes, etc. Each script entry should be serial numbered, so that, if the test ballots are entered in scripted order, the test ballots may be reconciled with the audit log entries.

Each machine, for this model, needs to be equipped with a set of placards, flags or other clearly visible signs to indicate what phase of testing it has reached. Brightly colored paper flags that could be hung from the magnetic clamp built into one of the side panels of the voting booth would work, for example. Consider a green flag meaning "open for test balloting", a white flag meaning "ready to close polls", a yellow flag to indicate "results being printed", and an orange flag indicating "discrepancy". The flag meaning should be indicated on the flag in large print, with small print giving instructions for what needs to be done next on this machine. An unflagged machine is either one where the testing has yet to begin or one where the testing has ended. In the former case, the machine will be equipped with a PEB and unused test documents, while in the latter case, the PEB will have been taken to the vote collection center.

Startup: One or more startup teams should proceed from machine to machine, verifying that the correct script folder is at each machine by comparing serial numbers, verifying that the correct script is in the folder, turning on the machine, and starting the zero tape printing. As soon as all machines are started, the startup teams are disbanded and their members assigned to other jobs.

Open Polls: One or more poll-opening teams, on finding a machine where the zero tape has completely printed, should verify that the serial number on the tape matches the serial number on the paperwork, and that all totals are zero. If these conditions are met, this team opens the polls on the machine and marks it with the agreed-on sign indicating that it is ready for voting. Once this is done on all machines, the poll-opening teams may be disbanded and their members assigned to other tasks, but some of these teams or team members may serve as roving supervisors for the testing.

Test Voting: Test voting teams move from voting machine to voting machine, looking for a machine that is open for voting (as indicated by the green placard or flag). On finding a machine open for voting, if the test script has been entirely voted, the team changes the sign on the machine to one indicating that the machine is ready to be closed.

If there are test votes remaining to be cast, the testing team enters and verifies the lowest numbered test ballot in the script. This could be done as it was on August 9, with one team member acting as voter and the other acting as scribe, or it could be done as suggested in the previous section, entering the test ballot, crossing it off from the script, and moving on to another machine, leaving the unrecorded test ballot to be recorded by a different team, with a blank absentee ballot of the appropriate format left lying across the screen but not covering the red button, as a signal that the machine is ready for use by the vote recording team.

If the script takes the form of pre-voted absentee ballots, as we did in on August 13, then, on completing entering and verifying the ballot, the test voting team should cast the ballot by pressing the red button and then enter, on the test ballot., the time at which it was voted, so that the test ballot can be reconciled with the event log, before putting the voted ballot in the envelope.

Note that, in this proposal, test voting teams move from machine to machine, never casting two consecutive ballots on the same machine. This avoids familiarity, forcing them to read the ballot

on each machine, and it encourages regular role changing between voter and scribe. Also note that, unlike the August 13 test, the test voting teams can begin work immediately when any voting machine is open for voting.

Vote Recording: If vote recording is done separately from vote entry, as suggested in the previous section, when a test voting team arrives at a machine containing a blank absentee ballot lying across the screen, they verify that this ballot is of the correct style for the votes that are on the review screen, then transcribe the votes from the review screen to the ballot, cast the electronic ballot, record the time of day on the paper ballot (to allow reconciliation with the event log), and deposit the paper ballot in the envelope before moving on.

Teams can be dedicated for vote recording, or the same team can either record votes or cast votes, depending on whether the next machine they encounter is ready for a new test ballot or is ready to have a vote recorded.

Poll Closing: On finding a machine marked with the sign indicating that its script has been completely voted, a poll-closing team should close the polls and start printing the precinct totals tape. While this tape is being printed, the agreed on sign should be posted to indicate that the results tape has begun to print.

Results Collection: On finding a machine where the results-printing sign is up, where the printer is stopped, a poll-closing team should investigate. If the results tape has printed to the end, the tape should be torn from the printer, the PEB should be put in the results envelope, and the seals on the machine should be cut so that the compact flash card or cards can be put in the results envelope. Note that the results tape and zero tape are on one continuous piece of paper at this point.

The poll-closing team should then compare the results from the printout with the predicted results from the script. If the results match the predictions from the script, the poll closing team should take down the status sign, put the results tape in the envelope, fill in the time-of-day at which the test was completed, put the remaining documents for the test in the envelope and immediately deliver the envelope to the test vote collection center.

If there is a discrepancy between the test plan and the result of the test, the poll-closing team should document the nature of the discrepancy and then mark that machine as having a discrepancy using the agreed on sign, put the remaining documents for the test in the envelope and immediately deliver the envelope to the discrepancy evaluation center.

Discrepancy Evaluation: Where a discrepancy is found between the predictions of the test script and the results tape, a discrepancy evaluation team should investigate. Such a team requires a computer for reading data out of the compact flash cards, probably a technician from the tabulating center and auditors, and it should operate under the close supervision of the canvassing board.

This team should take the results envelope and, using a computer, extract the event log from the compact flash card extracted from the voting machine and compare the record of events in the event log with the records of the test, including the times of day at which each test ballot was cast. This should quickly determine if any test ballots were not cast or if extra ballots were cast that were not part of the script.

If there is a video record of the vote or of the verification screens on that machine, or if there is an independent record of the votes cast (for example, if vote-recording teams and test-voting teams functioned separately), the discrepancy evaluation team should examine these records to further explore what happened.

If the discrepancy evaluation team identifies a specific error in the execution of the test script, they should clear the PEB for the machine in question and authorize the repetition of the test on that machine. A supervisory team should then prepare new paperwork (a new copy of the script, for example) for the machine while the machine is cleared, tested, and re-opened for voting, and then the appropriate sign should be put up on that machine to indicate that it is open for testing.

6) The Event Log Bug

The following message appeared in the event logs from the August 13 test, as extracted from the compact flash cards (abridged):

```
RUN DATE:08/14/04 02:37 PM                                ELECTION ID: 040831PR
Votronic PEB#   Type   Date       Time       Event
5115457 125934 SUP    08/11/2004 17:30:24   01 Terminal clear and test
          121061 SUP    08/13/2004 08:10:16   36 Low battery lockout
          08/13/2004 08:34:03   36 Low battery lockout
          08/13/2004 09:47:51   04 Enter service menu
```

Here is the same data, extracted from the serial port the next day (also abridged):

```
RUN DATE:08/14/04 12:25 PM                                ELECTION ID: 040831PR
Votronic PEB#   Type   Date       Time       Event
5115457 125934 SUP    08/11/2004 17:30:24   01 Terminal clear and test
          121061 SUP    08/13/2004 08:10:16   36 Low battery lockout
          08/13/2004 08:34:03   36 Low battery lockout
          08/13/2004 09:47:51   04 Enter service menu
```

This low battery message is, I believe, the message that would have been misrecorded in the internal memory of the iVotronic prior to the software updates of this summer. The consequence of this misrecording would have been the replacement of the terminal clear and test message immediately preceding the low battery message by a nonsense message in the data extracted by the serial port. In the data extracted via the compact flash card, the machine serial number 5115457 would have been corrupted. The fact that the records extracted by these two paths were identical and uncorrupted supports the claim that the event-log error from last spring has been fixed.

There was one other low battery report in the event logs that I collected from the tests (also abridged):

```
RUN DATE:08/14/04 12:25 PM                                ELECTION ID: 040831PR
Votronic PEB#   Type   Date       Time       Event
5111490 131401 SUP    08/08/2004 10:09:12   01 Terminal clear and test
          08/13/2004 17:56:45   12 Audit upload
          08/13/2004 17:57:05   36 Low battery lockout
          08/13/2004 18:00:27   14 Print Precinct results
```

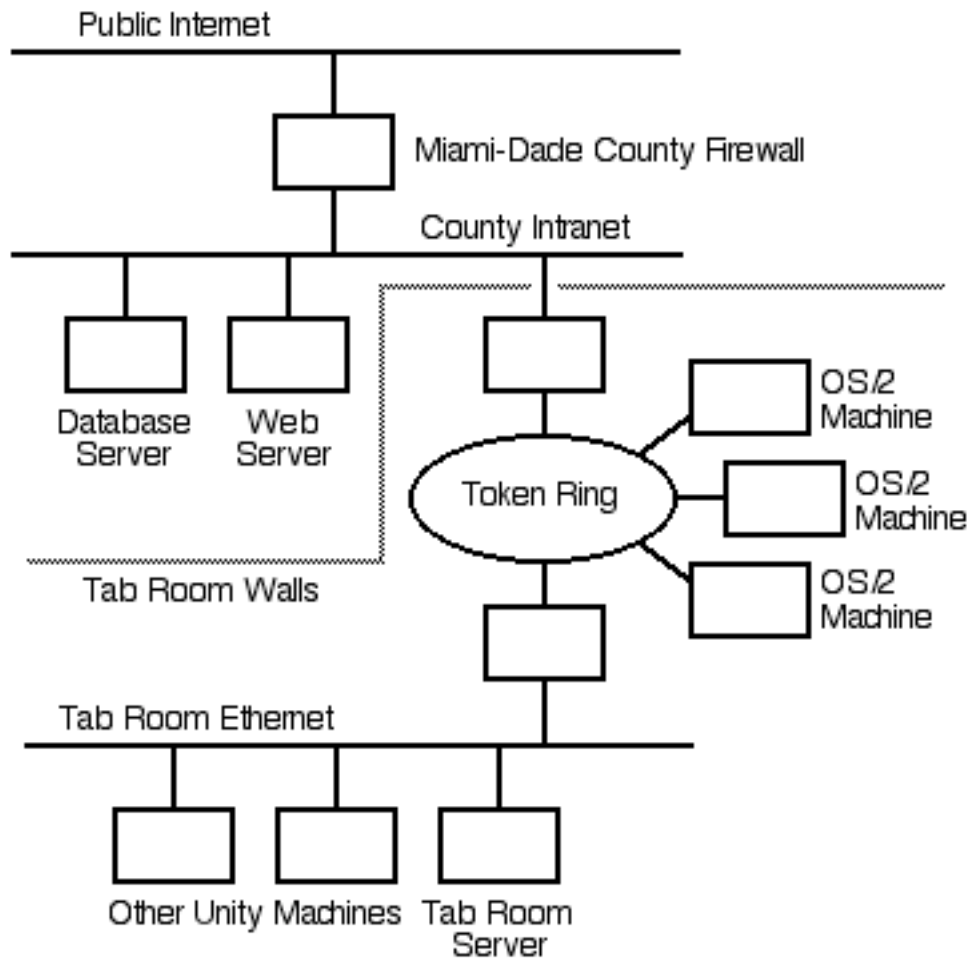
This message appears only in the data extracted via the serial port because the low battery condition occurred after the data had already been copied to the compact flash card, as recorded in the audit upload event. Prior to the software updates this summer, the low battery message and the audit upload message would have been combined to create a nonsense entry in the event log extracted by the serial port.

7) Security

While in Miami for the pre-election tests, I investigated one large security issue, that of the safety of the election management system in the face of potential threats from outside the elections office. I also have comments on the physical security of the ports on the back of the iVotronic, the policies surrounding the tabulating center, and the security of the compact flash cards.

Election management system security has been subject to significant criticism, for example, in the SAIC report commissioned by the State of Maryland last fall. There, it was suggested that all connections between the Internet and the election management system should be forbidden. The central problem is that we want a guarantee that nothing but authorized software and genuine election data ever enters the the election management system, while at the same time, we want fast and easy export of election results from these same computers.

In Miami-Dade County, there is an electronic connection from the election management system in the tab room to the outside world. This connection is a potential avenue of attack, and it is worthy of strong oversight. This connection is complex and obscure, sufficiently so that it should be a challenge to any hacker attempting to break through it, and it is entirely possible, even likely, that this connection is actually secure. Here is a figure that summarizes this:



The official election results are stored on the server within the tab room, the secure vote tabulating center. Various machines running the Unity election management system within the

tab room are used to build the databases on this server, using data downloaded by modem, data entered from compact-flash cards, and data from absentee ballot processing. All of these machines communicate using Internet protocols over a local network, the tab room Ethernet.

Export of data from the tab room (the vote tabulating center) is overseen by an applications programs running on the cluster of OS/2 machines, interconnected by their own token-ring network. OS/2 is an old but high quality operating system developed by IBM to compete with Windows. The machines on this small network communicate using the NetBIOS protocols, a protocol suite that is incompatible with the more common Internet protocols. The only data that is communicated through these machines is that data that applications programs on these machines deliberately lift over the wall created by this incompatibility.

The data export applications, running in the tabulating center, send election results to a database server attached to the county intranet, and also to a cable-TV content generation server that feeds the county's TV network, not shown in the figure. The election results on the county's database server are available only on the county's intranet.

When a public inquiry comes in, through the county firewall, it comes to the web server; the firewall prevents access from outside the county intranet to either the database server or the gateway machine into the tab room. The web server, on receiving a request for election results, converts this to a database query, and then takes the results from the database server and converts them into a web page delivering the requested information.

With the above picture of legitimate use in mind, it is appropriate to ask what vulnerabilities this system may have to attack from a malicious computer user intent on corrupting or disrupting an election. First, this attacker must gain access to a county computer, either physical access, if the attacker is a county employee, for example, or by hacking into a computer that is visible through the county firewall from the public Internet. The county intranet is large enough that the threat from county employees should not be discounted, and firewalls are not perfect; neither of these threats can therefore be discounted.

Once the attacker has access to a county computer, the attacker must break through the gateway machines to the tab room. These computers can easily be powered down or physically disconnected from the county intranet except when an election is in progress or when pre-election testing is under way, so this threat can be easily avoided except for a few days during each election cycle. The fact that the OS/2 machines and the token-ring network that serve as the gateway to the tab room are quite different from most computers on the Internet will deter some attacks.

Nonetheless, there is a remote possibility that an adversary could find a weakness of the machines that serve as a gateway to the tab room, and exploiting this, somehow find a weakness in the servers within the tab room during an election. While this possibility is remote, it is sufficient to justify defensive actions. The current procedures used in Miami-Dade County include, as one of the final steps in canvassing an election, the checking of the reports generated by the election management system against the paper tapes printed at each precinct when the polls were closed. This defense should be sufficient to detect and compensate for any attack on the tabulating center through its network connection to the outside world.

While I believe that Miami-Dade County is adequately defended, I am concerned by two issues: First, I worry about the ad-hoc way in which Miami's defenses have evolved to their current state. Evolutionary development is inevitable in large computer systems, since they are rarely put in place all at one time, but rather, they are replaced and expanded on a component-by-component basis. On the other hand, to rely on a somewhat elderly (but still robust) collection of OS/2 machines running a mix of homegrown applications and software dating back to the era of

punched-card voting may not be the best thing, particularly when it appears that this system is largely the child of one gifted system developer.

Second, with 3,142 counties in the United States, most with nowhere near the resources available to Miami-Dade County, I wonder how many have far less secure internet connections to their election management system and how many have canvassing procedures that involve double-checking the electronic results against the paper record from the precinct. It would be far better to rely on standard well-documented security models for isolating the tabulating center from the outside world instead of relying on local ingenuity to construct appropriate barriers, and it would be much better to rely on industry-standard software for exporting data across this barrier than to rely on home grown applications. I have suggested an outline of a reasonable model for this in my recommendations to ES&S, in the hopes that they can support such a model in the future.

The rear of the iVotronic has 4 sockets, one for power, one for serial data, and two (stacked one over the other) for the compact flash cards. There is a sliding door that can be closed over these sockets, and the machine has provisions for a security seal that can be used to prevent the door from being opened all the way, protecting the compact flash cards inside the machine.

The Compuware report for the state of Ohio complained that there was no provision to attach a second security seal to prevent access to the serial port. It is quite possible that the machine is not vulnerable to attack through this port, but in the best case, proof of this would require a meticulous audit of the iVotronic firmware, and in the worst case, such an audit may be inconclusive. Because of this, ES&S should provide for a second seal in future models of this machine.

The actual seals used are as important to the integrity of the system as the provisions for sealing the machines. During the August 13 tests, I found one seal on the ground in the testing area, and later, while the compact flash cards were being collected, the team collecting these cards found a machine with a missing seal. I do not know that the seal I found matched the seal that was missing, but this incident raises questions about the seals being used in Miami.

The seals used appear to be Tug-Tight seals, a widely available brand, sold by U-Line for \$68 per thousand. Very similar seals are available with custom printing from a number of vendors, such as American Casting & Manufacturing. In general, these lightweight plastic seals are designed to be attached with no tools and removed with no tools. They are excellent for use in contexts where accidental removal is unlikely or where accidental removal has no serious consequences. For example, when used to seal fire extinguishers, a common use, the only negative consequence of accidental seal breakage is that the next time the fire extinguishers are checked, those with accidentally broken seals will be unnecessarily taken away for recharging.

In contrast, the seals on the rear panel of an iVotronic are exposed to possible tampering by every voter, and if even one seal is accidentally broken in a tightly contested election, it could raise questions about the integrity of the election. Therefore, I recommend use of seals that are unlikely to be broken except by someone setting out to deliberately break them.

Also, with so many seal vendors offering custom printing on their seals, I strongly suspect that there might be one vendor of seals similar to the ones used in Miami that would be willing to print seals with a particular series of numbers on demand. If I know the seal number of a machine, it is quite possible that, in a matter of a few days, I could get a very similar seal with the same number made commercially, although it is likely I would have to order a thousand consecutively numbered seals that include the number I want.

The shape of the seal-holes on the iVotronic suggests that this machine was designed to be sealed by a cable-tie. Cable ties can be applied by hand and are extremely difficult to remove without using a wire cutter. American Casting & Manufacturing and Bay State Cable Ties, among several others, make consecutively numbered printed cable ties with custom printing. I strongly recommend getting custom printed seals that say something like "DADE COUNTY" on them (a 12 character limit is common), or, if available, the official Miami-Dade County seal, in the hopes that most printers will not print seals with this text or image for users other than Miami-Dade County.

Tab room security is also a matter of concern. Several people in the tab room during the pre-election testing were not county employees, and several others were temporary employees -- myself included. I never received any lecture on tab room security, and it was only the next day, during post-test analysis of the contents of the compact flash cards, that the members of the Miami-Dade Election Reform Coalition learned that their access to the tab room was very unusual.

There are reasons to be particularly concerned about the role of vendor employees in the tab room. It is clearly within reason to have on-site assistance from the vendor, including service-contract assistance to deal with the routine mechanical problems that are expected from machinery such as is used to tabulate absentee ballots, and to help with unexpected software issues such as the problem with releasing the interlock when there were more test ballots than registered voters in some precincts during the August 13 tests. On the other hand, I am concerned when vendor employees are allowed to operate election machinery without close supervision from regular elections department employees, particularly when they are allowed to touch ballots or to work with the election management system.

I would have been happier if all temporary employees and visitors to the tab room had received a lecture on tab room security -- including restrictions on the use of computer media in the room, since import of software, that might include computer viruses or other threats, needs to be very carefully monitored. Laptop computers, floppy disks, and even cameras with compact flash cards should be viewed with suspicion and closely monitored. It might even have been appropriate to ask each visitor to sign an agreement to abide by these security rules.

I would also prefer it if all temporary employees, contractors and visitors present in the tab room were required sign in, sign out, and openly display easily observable badges that allowed observers on the other side of the glass wall to easily understand who was there in what role.

Compact flash cards are commonplace today, and the compact flash cards used in the iVotronic are only weakly protected, according to the Compuware report commissioned by the State of Ohio. Because of this, it is essential that election officials, including precinct officials, view any manipulation of compact flash cards anywhere near the voting machines with considerable suspicion.

I was gratified, when I pulled the card from my camera and waved it near a voting machine, to have an employee of Miami-Dade County descend on me and sternly tell me not to do that! This is exactly the right attitude to take, and it is an attitude that should continue even when ES&S augments the security of election data with cryptographic technology.

Given the weak electronic security of the contents of the compact flash cards in the iVotronic, the county gains a small measure of security by using an uncommon type of compact flash cards, industrial-grade cards that are easily distinguished from the cards that are commonly available at retail. Even the county's habit of tagging each card with a piece of adhesive tape offers an additional, if small, level of security.

8) Absentee Ballot Counting

One test I performed in conjunction with the pre-election testing on August 13 was a test of the quality of the calibration of the ES&S model 650 absentee ballot scanners. This test is one I have routinely done as part of state certification testing for mark-sense tabulating machines brought before the state of Iowa, and my visit to Miami provided an opportunity to compare the calibration of one of these machines, in use in the field, with the calibration of the machines as they are brought forward to the states.

I also feel that, as a general rule, counties should not rely on the vendor's calibration of voting machines, but should routinely verify that the calibration matches an intuitive model of what marks ought to be counted as votes and what marks ought not be counted as votes.

For this test, I used a sequence of 10 ballots, "NON STYLE :0222" "Typ:03 Seq 0222 Spl:01" that were hand marked using pencil (number 2 soft lead) and two colors of pen (a purple fine-point felt pen, Kuretake ZIG Millennium .5 millimeter, and a blue roller-ball pen, Pilot Precise V5 extra fine). The tests included tests for uniformity of calibration across the three columns of the ballot, sensitivity to erasure, and reading of nonstandard marks such as check-marks, X marks and circles around the voting target.

The basic sensitivity tests made with the following series of marks, using pencil to vote for Theresa M. Pooler (column 1), Valerie R. MANNO (column 2) and YES/SI/WI (column 3), using a purple fine-point felt-tip pen to vote for Barbara ARECES, and using a blue roller-ball pen to vote for William L. Thomas:

	1	2	3	4	5	6	7	8	9	10
Pencil	+	+	+	+	+	-	-	-	-	-
Purple	+	+	+	+	+	-	-	-	-	-
Blue	+	+	+	+	+	+	-	-	-	-

A plus above indicates that the mark was counted as a vote, while a minus indicates that the mark was not counted. The quality of the images above is somewhat degraded because they are taken from photocopies of the original test ballots, which are included in the pre-election testing archive. Generally, mark number 5 was the faintest single stroke I could draw with that pencil or pen spanning the entire voting target, while marks 6 and higher were successively shorter. The three pencil tests in different columns of the ballot showed identical sensitivity.

Miami-Dade County instructs voters to use an X through their selection to make corrections, forcing an overvote so that the canvassing board will see corrected ballots. With absentee ballots, it is inevitable that some voters who have voted using pencil will attempt to make their corrections using an eraser instead of following these instructions. Therefore, I tested the ability of the scanner to process erasures, voting for Karin A. Brown:


















	1	2	3	4	5	6	7	8	9	10
Erasure	+	+	+	+	-	-	-	-	-	-

A plus above indicates that the mark was counted as a vote, while a minus indicates that the mark was not counted. The image quality above is degraded because these images were obtained from photocopies of the originals. Ballot number one was a dark pencil mark, while

ballots 2 to 10 were erased, with a single eraser stroke on ballot 2, two eraser strokes on ballot 3, three eraser strokes on ballot 4, and so on. There was a faint grey film visible in the oval on all erased ballots, although it was progressively fainter with each additional eraser stroke. What these images show, effectively, is that the mark sense scanner and the photocopier had very similar sensitivities. Where the photocopier picked up residual marks despite the attempted erasure, the scanner counted a vote, and where the photocopier saw no mark, neither did the scanner.



My conclusion from these experiments is that the ES&S model 650 scanner I tested was calibrated very close to what I would hope for. Conscientiously made marks were always counted, no matter what pen or pencil was used. A single firmly-made pen or pencil stroke across the width of the voting target was always counted. The scanner ignored erasures that were done with a reasonable degree of care, and it ignored small dots, typical of both flyspecks, printing defects and hesitation marks (pencil or pen marks made by the voter unintentionally resting the tip of their pencil or pen on the paper while reading down the ballot).

The final tests revolved around nonstandard marks on the ballot, checks, x-marks and circles. The ten experiments for each marking are shown below, using votes for Carlos A. ALVAREZ, Ada POZO REVILLA, Rose L. EVANS-COLEMAN and Peter ADRIEN (in order, from top to bottom):

																																																																																																																																																																																																																																																																																																																																														
--	---	---	---	---	---	---	---	---	--	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

instructions have several defects that experience shows lead to a small number of voters misreading the instructions. Here is the English version of the instructions used in Miami:

**TO VOTE, COMPLETELY FILL IN
THE OVAL  NEXT TO YOUR
CHOICE.**





**Use black or blue ball point pen.
If you make a mistake, cross
through the mistake  and fill in
the oval  next to your correct
choice.**

These instructions do not show the oval before it is filled in. For voters with sufficiently poor eyesight, this appears to have led a small but significant number of voters in the 2000 election to find places on the ballot that resembled ovals and fill them in. This problem was compounded by the second problem with these instructions.

These instructions do not show the oval in relation to the text in the same way the oval relates to the candidate name. The Spanish version is at the worst, actually showing one of the example ovals all the way to the right of the text, therefore predisposing voters to look for something to fill in on the right of the candidate name. On the Miami-Dade ballot, candidate position numbers are on the right of the candidate name, where they could serve as tempting false voting targets.

In 2000, similar faults in the design of the instruction used on the Global/Diebold mark-sense ballots predisposed a significant number of voters to darken the (DEM) or (REP) used by some counties. These parenthesized party abbreviations were the same size as the voting targets and were placed to the right of candidate names while the voting target was to the left.

I suggest that this problem could be solved by rewording the instructions as follows; this alternate set of instructions takes no more space on the ballot than the original (at least, in English), and it avoids the possibility of misdirecting voters to expect voting targets in any place but where they are, relative to the text on the ballot, while also showing an un-voted target:

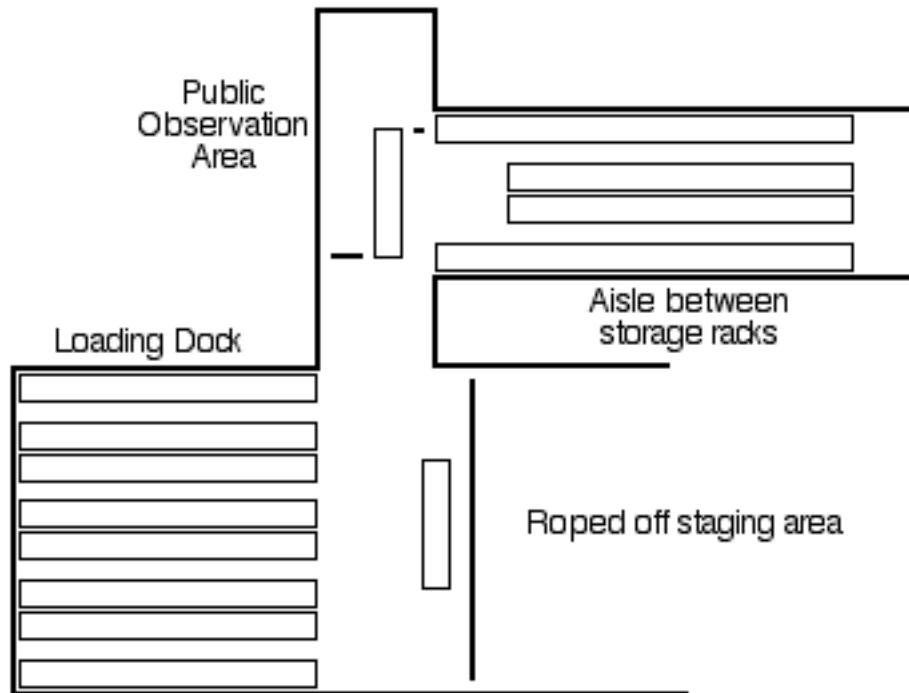
 **TO VOTE, FILL IN THE OVAL
 NEXT TO YOUR CHOICE.
Fill it completely using black
or blue ball point pen.
If you make a mistake, cross
 it out and fill in the oval
 beside the choice you prefer.**

Any change to the instructions, must, of course, face close scrutiny. Even though the original instructions were almost certainly never subject to any serious experimental testing, they have seen modifications over the years in response to state and county suggestions, and any additional change will, quite rightly, be questioned. Certainly, the form of instruction should not be changed without consulting the state election office! It may already be too late to make changes for this fall.

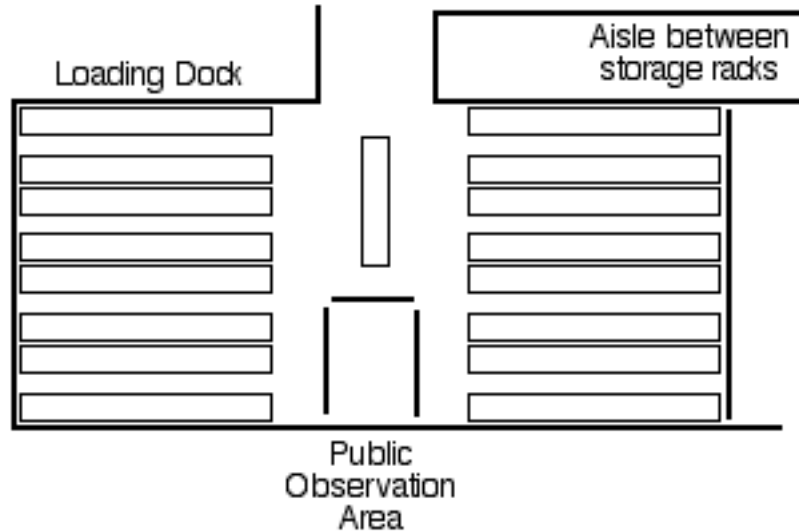
10) Openness and Observability

A public pre-election test of the election machinery serves no purpose unless representatives of the public can observe the test. As mentioned in part 5 above, testing that lasts until after midnight is not easy to observe. On the other hand, public observation can get too close, for example, when members of the public are able to roam freely, through the testing area without close supervision. This occurred, to some extent, on August 13 because of the physical layout of the testing area.

The use of warehouse space is tightly constrained! The use of the loading dock area for the bulk of the testing was an excellent choice, but the use of the aisle space between two storage racks posed problems, creating two separate testing areas that could not be observed at the same time, either by the canvassing board or by representatives of the public. The following diagram shows, in a very schematic way, how the available space was used:



Because the public, and the press, could not effectively observe the process from the area where they were initially to be confined, they escaped, and the result was both a potential security problem and a minor, though at times annoying hindrance to the testing. Had there been a way to arrange the testing so that all of the machines under test could be observed from one area, the situation would have been far better. The following suggestion might have been possible, although it would require rearrangement of the stacks of iVotronics that were ready for trucking to the polls. Other arrangements might also have been workable.



The effort made by the county to explain the testing to the press and to members of the public who came to observe was quite good, for those parts of the testing that were conducted in the warehouse. On the other hand, as testing moved into the tab room, the attitude changed.

When testing moved to the tab room, part of the change in attitude can be attributed to the long day, the late hour, and a reaction to the loss of control earlier in the day, but some of it appeared to be systemic. Tabulating center procedures are far more technical and harder to explain than the procedures followed during the hands-on tests of voting machines. Furthermore, the glass partition between the public viewing area and the tab room, along with the one-way communications channel from the tab room to the viewing area completely changed the relationship between observers and election officials that had existed in the warehouse phases of the tests.

Where, in the warehouse, the observers had, if anything, too much access, I feel that, when testing moved to the tab room, the observers had too little. When election observers do not understand what they are observing, the right to observe has very little meaning.

When Geneva Switzerland moved to Internet voting, according to Michel Chevallier, Secrétaire adjoint, Chancellerie d'Etat de Genève, one of their biggest early mistakes was a failure to explain, to the observers, what they were seeing. To the observers, all of the activity in front of computer screens in the Geneva tabulating center might as well have been magic. Chevallier concluded that observers should be offered training in the use of the election management system so that they can understand what is being done.

I do not believe that most observers from parties, the press or the public would have the time to take training in the use of the election management systems in use in the United States, but at the very least, the instruction manuals for the tabulating machines and election management systems in the tab room should be available to observers, preferably on the internet so they can study them in advance. Even this is not enough; having read the manual does not necessarily give someone the knowledge to understand what someone on the opposite side of the tab room is doing unless the computer screen itself is visible to the observer.

Openness in election administration really does require explaining what is going on to the press and public who are observing the tab room. The public has every right to interpret a failure to explain as being suspicious, even in the near battlefield environment of an election. Therefore,

someone in the tab room ought to be regularly reporting to those who are observing, telling them what it is they are seeing, and taking questions. An open microphone in the tab room that simply broadcasts the hubbub of the room is not sufficient.

It was very difficult for observers to identify the cast of characters in the tab room. If this is true during testing, it is even more true and even more important during an election. Clear and visible badges indicating the status of each person in the tab room would have been helpful, as was discussed in Section 7. Numbered football jerseys with a playbill for the observers might be going too far, but only slightly.

Finally, something must be said about the privileged position of the Miami-Dade Election Reform Coalition. Their representatives were present in the tab room throughout the late phase of the pre-election testing; in a very real sense, this represents a potential breach of security. I say, potential breach because I believe that all of the members of the coalition who were present were people of very high integrity, but I believe that to institutionalize a class of privileged observers with special inside access sets a dangerous precedent. I would have preferred a system that was sufficiently open and observable that there was no need to create a class of privileged observers, but where, instead, all of those observing through the glass window would have had sufficient information about what was happening on the other side of the glass that they did not feel excluded or shut out by being on the other side of that glass wall.

During the post-test followup work the next day, however, it was not unreasonable for the very small number of observers present to be in the tab room with the tab-room staff and audit and management staff, although I would have been happier, as I pointed out in Section 7, if all of those who were not permanent employees of the elections division were given a security lecture before admission to the tab room.

11) Touch Screen Calibration

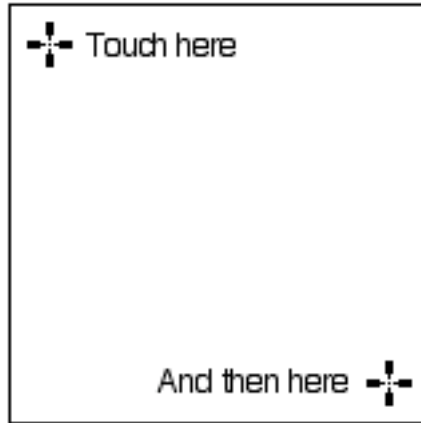
There have been several reports of calibration problems on touch screen voting systems. Here is a typical report of such a problem from the news media:

The shortcomings included "a calibration problem" in some machines that had voters selecting one candidate and the screen displaying another.

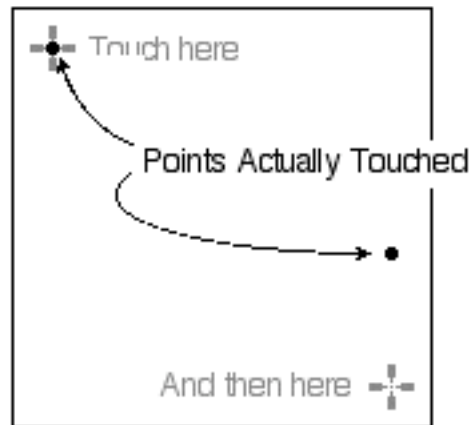
St Petersburg Times, Broward official apologizes for voting mess, Sept 20, 2002

In order to understand the touch screen calibration problem, it is important to understand that all touch screen computer systems involve two separate devices, one is a display screen, an output-only device, and the other is a transparent touch sensor, an input-only device. When these are properly aligned, touching a spot on the touch-sensor will be interpreted as being at the same location as the point on the display screen directly under this spot, but simple mechanical alignment is not sufficient to assure this for the most common touch-screen technologies. Therefore, most touch-screen computer systems include, as part of their routine setup procedures, a calibration step that involves touching selected spots on the screen in order to allow the computer to learn how the touch sensor is aligned relative to the display screen. If this calibration is done incorrectly, a touch to one spot on the screen will be interpreted as a touch to a different spot.

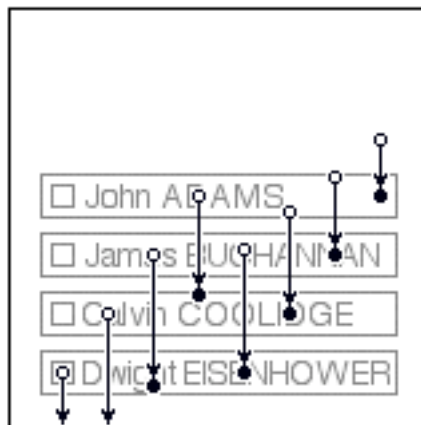
The touch-screen calibration routine typically presents two targets on the screen, in opposite corners. Sometimes a third target is added to help detect and reject deliberate or accidental miscalibration. If the user follows the touch-screen calibration directions precisely, touching the calibration targets at their exact centers, the computer will be able to interpret all later touches extremely accurately. The typical touch-screen calibration display resembles the following:



The worst-case allegations about touch screen calibration are that some touch screens have been deliberately miscalibrated in order to favor one candidate or another. Consider, for example, the consequences of following the calibration instructions as follows:



Having miscalibrated the screen this way, the consequences, for a voter, would be as follows, illustrated on a ballot showing John ADAMS, James BUCHANNAN, Calvin COOLIDGE and Dwight EISENHOWER, all listed near the bottom of the screen:



This miscalibration was done at the bottom of the screen, so the consequences are more severe there. The effect near the top of the screen will be much smaller. As illustrated above, touching COOLIDGE or EISENHOWER will not be recorded as votes, touching ADAMS will be recorded as a vote for COOLIDGE and touching BUCHANNAN will be recorded as a vote for EISENHOWER.

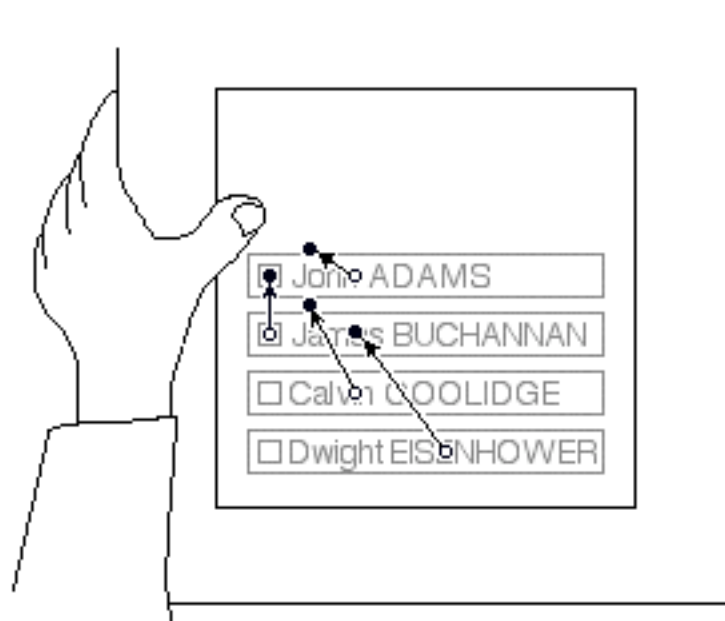
The problem with this theory of deliberate miscalibration is that the same miscalibrations that lead to voting problems will also throw off touch-screen use by polling place workers. Therefore, it is extremely unlikely that this kind of problem would go unnoticed. Furthermore, it is not straightforward to design a miscalibration that will throw votes in a deliberate direction, as opposed to merely making a machine difficult to use. As a result, I suspect that deliberate touch-screen miscalibration, if it has occurred at all, is very rare.

The possibility of accidental miscalibration is of far greater concern. Can an iVotronic (or other touch-screen voting system) become miscalibrated as a result of shock? Could a malicious voter vandalize a machine in a way that miscalibrates it?

In order to explore this, I experimented with an iVotronic that was not part of the pre-election testing, the demo machine in the lobby of the elections offices. It was very difficult to cause any perceptible calibration change by deforming the case at the edge of the screen. I conclude that there is an extremely remote possibility that shock to the voting machine case could throw off the calibration by denting the rim of the display screen so it presses against the touch-sensing surface. I suspect that there would be visible damage to the case if this did occur.

I also tried to simulate "vandalism" that could be intended to throw off the calibration of the display without being otherwise visible. One way to do this is to insert something between the edge of the case and the touch-sensing surface in order to apply pressure to the surface. Things wedged in this crack were quite visible, and furthermore, it was extremely difficult to adjust the wedge so that it had any effect on the touch-screen calibration. Furthermore, the calibration always returned to normal when the debris was removed.

A voter can accidentally simulate miscalibration by holding the edge of the machine with one hand while voting with the other. This is only possible for voters with hands large enough that the thumb of the idle hand rests on the screen. I found this easy, and I imagine that there are some voters who might not notice themselves doing this while they try to vote. When the iVotronic screen is touched with two fingers, the computer averages the positions of the two fingers. The effect of this in a voting context is illustrated below:



In this example, I would have difficulty voting for ADAMS, but I would probably succeed by touching his name a little low. An attempt to vote for BUCHANNAN would be interpreted as a vote for ADAMS. An attempt to vote for COOLDIGE might be difficult but could be likely to be interpreted as a vote for either ADAMS or BUCHANNAN, depending on whether I touched high or low on the COOLDIGE area, and an attempt to vote for EISENHOWER would be interpreted as a vote for BUCHANNAN. Of course, the moment I let go of the iVotronic with my idle hand, everything will return to normal, but I can imagine a voter becoming quite frustrated without ever thinking to let go.

The research required to identify whether this problem accounts for a significant fraction of the reports of miscalibrated touch-screen voting systems would be expensive, since these reports are rare, and this implies that a large number of experimental subjects would need to be observed in order to pin down this problem. Without such research, I believe it is inadvisable to suggest that reports of miscalibration were the result of this possibility.

It might be worthwhile to alert polling place workers, asking them to report if they observe voters holding the machine with their idle hand, and I strongly urge that routine incident reporting from polling places include reporting every voter complaint of the form "I tried to vote for X and it recorded it as a vote for Y." After hearing such a complaint, the polling place worker should check the calibration (easily done on an iVotronic); if the calibration appears good, the polling place worker should ask the voter if they were had held the machine with his or her idle hand while voting. If a significant fraction of these reports suggest that this was the case, the iVotronic case should be redesigned to include a ridge on each side of the screen to hold voter's thumbs back from the screen.

12) Recommendations to ES&S for Improvements to the iVotronic and Unity

During pre-election testing of the iVotronic in Miami Dade County on August 13, 2004, I observed several problems caused by aspects of the design of the iVotronic and Unity that would have materially simplified voting system testing, improved security, and in some case, improved the use of the iVotronic in the polling place. I do not expect any of these to be changed in time for the general election in November, but they are worth putting in the queue of ideas to consider for future system releases:

Ballot style selection -- The touch boxes on the screen where the polling place worker touches to select the ballot style are very small and spaced very close together compared to the touch boxes used for voting. My fingers are big. As a result, when selecting ballot styles, during testing, I selected the wrong style several times.

These errors led to cancelled ballots in some cases, but where test ballots of the style accidentally selected needed to be cast anyway, the errors were "covered up" by simply voting the style that had been accidentally selected. This "coverup" has no impact on the validity of the pre-election test, but the fact that we could and did do this means that no retrospective audit of the pre-election test event logs will disclose the real frequency of this error.

There may be no need for the visual glitz in the screens and menus intended for the polling-place worker, but there is good reason to make the same allowances for finger-size and dexterity that apply to screen displays seen by the voter. I recommend that the touch boxes for ballot style selection be spread out so that there is as much space between the different touch boxes as there is between candidate names on a ballot.

Error message presentation -- When a voter attempts to select more than the allowed number of candidates in a race, the iVotronic erases the entire screen, presents an error message saying

something like "please deselect a candidate before you make an additional selection" and then returns you to the screen for the race in question. This full screen erase, when I encounter it, destroys my sense of context thoroughly enough that I'm disoriented by it and disoriented again when I return to the original context.

Ben Bederson at the University of Maryland is doing research on this kind of issue, he may have some good advice to offer. My suggestion, based only on intuition, but based on 30 years of experience using various GUIs, is that the ideal error-popup would be more in the form of a dialog balloon growing from the touch that could not be accepted, and the balloon should clear either when the voter touches the OK button in the balloon or when the voter touches an already selected candidate, clearing the conflict that caused the balloon to appear in the first place. This way, the error message does not destroy the user's sense of context.

Review screen for multi-candidate elections -- the review screens for single-candidate races are very easy to follow, but when there is a multi-candidate race, the review screen only shows if you have selected fewer than the allowed number or all of the allowed number of candidates. In testing, this means you have to break stride, descending to the page where you would change your vote, if you wanted to, in order to verify that the test ballot has been properly tested. Breaking stride like this causes loss of context and slows the test.

For a voter, I imagine that many voters would not quickly grasp how to review their selections in a multi-candidate race, and so, I suspect that many voters who make errors as they vote in multi-candidate races never see them on the review screens and therefore never correct them.

It would be far better, I imagine, if the selected candidates were all displayed in the review screen for multi-candidate races. Space allocation for this on the review screen can be static. Reserve enough space to display the maximum allowed number of selections, and if the voter has selected fewer than this maximum, show "no candidate selected" for the later slots.

Machine hangs -- Once, while working through our stack of test ballots, my test partner and I managed to hang the iVotronic we were working on. In a polling place, this would have been called a pollworker error. Apparently, one of us pulled the PEB out of the machine at just the wrong time during ballot-style selection, and the machine became non-responsive, requiring technical assistance to get it restarted. Most of the time, if the PEB was removed prematurely, the machine reverted to its initial state, so you could start over, but apparently, there are still windows of vulnerability when removing the PEB puts the machine into odd states.

Obviously, these windows of vulnerability should be repaired. In addition, it would be useful if, when a PEB is in the machine, a display said something like "Do not remove the PEB" or "You may now remove the PEB". The former display should remain until the latter is true.

Security seals -- There is no way to seal the sliding back door so it covers the serial port and no way to seal the machine to prevent PEB insertion. The next time a new die is sunk for the plastic case of the iVotronic, I recommend adding new sets of seal attachment lugs so that seals can be applied to prevent access to the serial port and to prevent insertion of the PEB.

I would recommend the following procedure for using the second set of seals on the back door: Before shipping machines to the polling places, apply seals to both sets of lugs, so that only the power port is exposed. At the precinct, cut the second seal on only the machine to which the printer pack is to be attached and record its number in the precinct paperwork. Leave the printer pack attached to that machine, or, after printing the zero's tape, re-seal the machine and write down the new seal number. After the polls close, first consolidate all the data into the master PEB, then cut the seals and extract the compact flash cards, and finally print the poll-closing report. Extracting the compact flash cards as soon as possible is important.

The new pair of seal-lugs on the PEB docking bay would allow for a seal blocking insertion of the PEB in the iVotronic, thus offering some physical security against turning on the machine between the time the iVotronic is configured for a particular precinct and the time this seal is cut in order to open the polls. In Miami, where the zero tape is being printed the night before the election, a new seal could be inserted here to prevent insertion of the PEB between the time the zero tape is printed and the time the polls open in the morning.

Logging ballot cancellation -- It would be very desirable if, every time a ballot was cancelled on the iVotronic, a record was retained of why the ballot was cancelled. There is nothing wrong with retaining such a record in written form, but it would be far more valuable if the reason was included in the event log, so that computerized analysis could uncover patterns of cancellations. I suggest that a reason code be required to cancel a ballot. Some obvious reasons are:

- Wrong ballot style selected (probably the most common reason)
- Voter fled -- in a context where the rules require cancellation
- Machine accidentally enabled for voting when no voter present
- Apparent machine problem requires clearing ballot (so voter can use different machine)
- Apparent voter problem requires clearing ballot (starting over fresh better than editing)

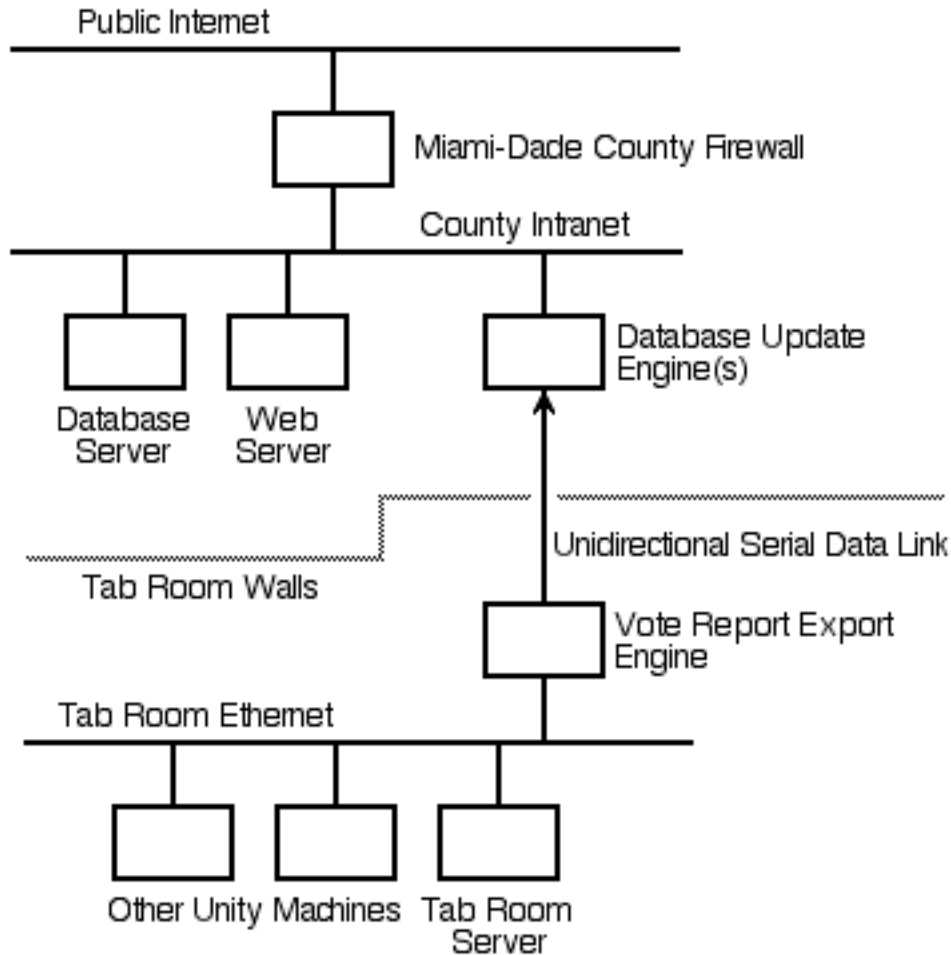
These should probably be discussed with election officials in a variety of jurisdictions before pinning them down in a new software release.

Election management system data export -- As discussed in the security recommendations in Section 7, it would be very desirable all voting system vendors would agree on a standard data export model so that counties would not have to rely on their own ingenuity to establish secure one-way connections to public networks. I recommend the following framework.

Use a very simple one-way data transfer connection, something physically simple such as a modified serial data cable that allows data flow in only one direction. Technically, this would be a null-modem cable with only signal ground, protective ground and TxD (transmit data) connections to the machine inside the tab room, and only ground and RxD (receive data) connections to the machine outside the tab room. Ideally, this cable should have exposed wires where it plugs into the serial port on the vote report export engine, although a transparent plastic shell on the connector to the serial port might suffice,

More technologically complex connections might be considered, such as a one-way connection using optical technology, but the important characteristic required here is that the connection be genuinely one-way and based on trivial enough technology that this fact can be easily verified. Modern high performance inter-machine connections do not meet this requirement! All allow bidirectional data transfer and are very difficult to convincingly secure for one-way traffic only.

The vote report export engine should, ideally, be a standard voting system component available from the voting system vendor. It would regularly scan the database, in the tab room server, of results approved for public release, converting them to a standard format resembling a wire-service stock-ticker data stream. This data stream would flow out over the serial data link, where security auditors could easily attach a wire-tap, for example, a PC configured as a terminal emulator, to verify what is passing over the connection out of the tabulating center.



The database update engine or engines should monitor the data flowing into them from the serial data link, constructing database update requests from each record they see and using these to update the public database or databases of election results. Assuming that the data format flowing over the report export engine is sufficiently simple and in a standardized format, this software should be easily maintained and easily customized to work with different output presentation formats.

It is important to note that the database update engine may miss an occasional report. When this occurs, the update engine has no way to request the retransmission of the missed report because there is no reverse channel back into the tabulation center. This requires that the report export data stream contain repeated copies of the data, and that each data item in this stream be self contained and not rely on the assumption that all previous items were correctly received.

Overriding sanity checks in Unity -- It was good news, during the August 13 pre-election tests, to find that Unity includes sanity checks on data, so that it detects and flags for human attention precincts where the number of votes being reported exceeds the number of registered voters. This caused problems during pre-election testing because the number of votes cast in the test was determined by the desired test coverage and not by voter registration, but despite this, the presence of these interlocks is welcome.

I am concerned, however, that it was very difficult to release these interlocks. There are actually legitimate reasons for the number of voters to exceed the registration in a precinct. The most

notable is that the only useful registration number for this interlock is the number of active registered voters. If an unusual number of inactive voters are mobilized by an election, turnout could well exceed the number of active voters that was used to set the interlock. Anytime an interlock condition in Unity is overridden, it should be extremely public, and the override should attract the attention of observers at the time it is done and the attention of post-election auditors. If I had my way, a klaxon would sound in the observer's room to warn that this was being done.

At the very least, Unity should allow overrides of interlock conditions, but this should require an authorization code, a name and a reason from the person requesting the override, and there should be a record of all this in the event log for the election management system.

Mark-sense ballot instructions -- The instructions for ES&S mark-sense ballots are fairly uniform across the country. The instructions used in Contra Costa California in the recall election of October 7, 2000 and the instruction used in Miami this August, have the same basic faults. The voting target is only illustrated in filled-in form, not in its unmarked form, the voting target is printed in the text in a location that fits the grammar, not so it is in the same relationship to the text as the real targets are to candidate names, and there are false targets to the right of each candidate name (ballot position numbers in Miami, party names in Contra Costa County). ES&S should suggest that all its customers to move to better instructions, as suggested in Section 9.