**Data Files Resulting from the Vote Analysis Code**

**2 November 2011**

Our current program(s) produce a number of different files with different data in them. This is intended to describe the data in the files.

At present, all files have file names beginning with the county (all lower case) and the date. The date is written YYYYMMDD so that it will sort appropriately. For example, the data for Beaufort County for the 2 November 2010 general election would have file names

beaufort\_20101102\_FILENAME

where the FILENAME part would be replaced with one of the file names mentioned below. We will use beaufort\_20101102 as our example file name prefix in what follows.

In the text that follows, sections in *italics* are special comments regarding problems that have been observed somewhere in the state but that can be detected from running this program on the audit data.

***The primary file containing information about errors, actions contrary to procedure, and such is the EXCEPTIONS file described below.***

1. **VOTE COUNTS**

The most obvious files are those that contain vote count summaries.

The original data file is the “155” file of the cast vote record. This has each vote recorded, with precinct name and number, iVotronic (“terminal”) serial number, ballot style number, candidate sequence number, and the candidate and contest data.

**beaufort\_20101102\_PctIvoBSContestCandidate.txt:**

This file contains the 155 data accumulated by ballot style:

Vote count

Precinct number

Terminal serial number

Ballot style number

Candidate sequence number

Candidate and contest textual data

In this, the “vote count” is the number of votes cast for the candidate within that precinct, terminal, and ballot style.

This is the first level of accumulation of vote counts, namely by ballot style within a precinct and terminal.

**beaufort\_20101102\_PctIvoContestCandidate.txt:**

This file contains the 155 data accumulated by precinct and terminal number:

Vote count

Precinct number

Terminal serial number

Candidate and contest textual data

In this, the “vote count” is the number of votes cast for the candidate within that precinct and terminal. This file is produced by summing the “PctIvoBSContestCandidate” data (the previous file above) across ballot style numbers.

This is the second level of accumulation of vote counts, namely within a precinct and terminal.

**beaufort\_20101102\_PctContestCandidate.txt:**

This file contains the 155 data accumulated by precinct:

Vote count

Precinct number

Candidate and contest textual data

In this, the “vote count” is the number of votes cast for the candidate within that precinct. This file is produced by summing the “PctIvoContestCandidate” (the previous file above) across terminals.

This is the third level of accumulation of vote counts, namely within a given precinct. These numbers should match the numbers reported by the SCSEC.

**beaufort\_20101102\_ContestCandidate.txt:**

This file contains the 155 data accumulated for each candidate and contest across the entire county:

Vote count

Candidate and contest textual data

In this, the “vote count” is the number of votes cast for the candidate within that county. This file is produced by summing the “PctContestCandidate” across all precincts in the county.

This is the final level of accumulation of vote counts, and these are the county-wide totals. These numbers should match the numbers reported by the SCSEC.

CAVEAT: The numbers in our file will include all votes in the 155 data. This may include provisional, etc., ballots that were deemed invalid, so there may be occasional slight discrepancies. These should be identifiable by comparing the precinct totals with the precinct numbers used for provisional, etc., ballots.

**Complications with vote counts.**

TOO FEW VOTES IN THE 155 COUNTS: The files described above come from counting the votes in the 155 file. If there are terminals whose Compact Flash memory cards have not been uploaded to the county-wide 155 file, then those votes cannot be counted and the counts in the files described above will be smaller than the counts obtained through the PEB accumulation.

TOO MANY VOTES IN THE 155 COUNTS: In contrast, if the memory cards have been collected from a terminal and the data uploaded to the 155 file, but that terminal did not have its votes counted in the county’s master count, then there will be more votes in these files than have been certified.

This usually happens for one of two reasons.

*If more than one PEB has been used to close terminals and collect vote totals, but only the one official PEB has had its votes accumulated at the county level, then there will be uncounted votes.*

*The other reason that we have seen vote totals exceed the certified totals is if there have been test votes loaded into the 155 file.* This is usually identifiable because the number of votes is small and because the votes are scattered across a number of precincts. We have not been able to determine exactly what causes test votes to appear in the master 155 file. One possibility is that the 155 file is not cleared after testing, and another is that a CF memory card used for testing is also used on election day, resulting in two sets of files uploaded to the county’s master 155 file. We see such test votes, for example, in Horry and Lexington counties.

1. **FILES REPORTING THE ELECTION PROCESS**

There are a number of files that report on the election process. These should be viewed largely as automatic checks on whether official procedure was followed. Major exceptions to the process are listed in a separate file, but this set of files can also be used to gain this information.

Most of these files are produced by reading the official SCSEC website results (referred to as the “detail” file because the files downloadable from the SCSEC website have that name) and the 152 event log file. The 152 file contains all the logged events with terminal serial number, PEB serial number (if known), date and time as recorded by the terminal, and the message code and expanded text version of the event log message.

**beaufort\_20101102\_PEBDetail.txt:**

This is primarily a list of the PEB serial numbers found in the 152 file. If all the CF cards have been collected from all the terminals, then this should be the complete list of PEBs used in the election. We sort these by whether or not the PEB has been used for closing (and thus appears in the 152 file) and whether or not the PEB has appeared in the 68A file as having vote totals uploaded to the county master. PEBs that appear in the 68A file have their upload times indicated. Sometimes there are multiple upload times.

We note that most but not all the PEBs have six-digit serial numbers. PEBs have been observed with three-digit serial numbers, and the seven-digit number 9999999 has been observed.

**beaufort\_20101102\_PctDetail.txt:**

This file contains the SCSEC detail data and the 152 and 155 data accumulated for each precinct. *This is the file that best indicates whether or not all the Compact Flash memory cards have been collected from the terminals and have had their data uploaded to the master file at the county.*

This file has two lists, with the same information in each. In the first half of the file the precincts are sorted alphabetically by name. In the second half of the file the precincts are sorted by number.

FOR EACH PRECINCT:

The first line has the precinct number and name

The precinct number

The precinct name

The number of voters in the precinct according to the SCSEC data

The number of votes cast in the election according to the SCSEC data

The turnout percentage (votes cast/voters) according to the SCSEC data

An indicator (“DATA MISSING” or “PCT MISSING”) if the 155 file cast vote count is not the same as the cast vote record count as certified by the SCSEC. This would indicate that some of the CF memory cards have not been collected from the terminals or that PEBs used to collect vote totals have not had those totals uploaded, so that those votes have gone uncounted.

The second line is a list of the terminals used in the precinct.

The third line is a list of the PEBs used to open and to close terminals in the pct.

It is to be noted that the absentee, failsafe, provisional, emergency, and similar precincts are likely always to be labeled as “missing” data. This is partly due to the fact that the paper absentee votes are not included in the 155 file, so the totals obtained from the 155 file for the absentee precincts are almost always going to be different from the certified counts.

It seems that in all counties except Beaufort and Charleston the precinct location are distinct and thus that a terminal can be associated with only one precinct. In Beaufort and Charleston there are cases in which more than one precinct votes in the same location and thus that terminals are not uniquely identifiable with a precinct. This distorts some of the statistics in the obvious ways, for example in that two precincts that both vote in the same location with the same terminals will show up with identical data in this file for the terminals and PEBs. There are also “errors” in reporting the data for Oconee and Spartanburg counties. In Oconee, there are four “Seneca” precincts but they are reported together in the SCSEC report, and in Spartanburg there seems also to be a precinct split that is hard to disambiguate when collecting and processing data. The numbers all add up, but the output files don’t necessarily state that fact without a small bit of further analysis.

**beaufort\_20101102\_IvoDetail.txt:**

This file contains the 152 and 155 data accumulated for each terminal.

The first entry contains two “Y/N” flags to indicate whether this terminal appears in the 152 file and in the 155 file, respectively. These entries should both be “Y” but sometimes are not. A “Y N” indicates that a terminal appears in the 152 but not in the 155. This should only happen if there were no votes cast on the terminal. A “N Y” indicates that the terminal does not appear in the 152 file but does appear in the 155 file. This seems to happen, for example, with the test vote terminals, but it also happens occasionally for other terminals.

The next entry is the terminal serial number.

The next two entries, in parentheses, are the PEB serial number and the date and time the terminal was opened, and the PEB serial number and the date and time the terminal was closed. This data comes from the 152 file.

The next two entries are the vote cast counts for that terminal from the 152 file and from the 155 file, respectively. One of these will show up as “XXXXXX” if the terminal does not appear in both files.

The next entry is the list of ballot style numbers used in that terminal.

The next entry is the list of precincts in which that terminal was used.

The final entry is the list of dates and times when the CF memory card for that terminal was uploaded to the county master files. This comes from the 68A file.

**beaufort\_20101102\_EventLogText.txt:**

This file contains the code numbers and the texts of the actual list of events that appeared in the 152 file.

**beaufort\_20101102\_EventLogs.txt:**

This file contains the list of the event log messages, in the order in which they appear in the 152 file, but with only the message codes present and not the expanded text of the messages that correspond to the codes.

**beaufort\_20101102\_EventLogCounts.txt:**

This file contains the event log messages, but summarized with the frequency count for each message in each terminal.

**beaufort\_20101102\_EventLogCombos.txt:**

This file contains the frequency counts of the pairs, triples, and quadruples of successive event log messages for the given county. In determining pairs, triples, and quadruples, we exclude all pairs, triples, and quadruples in which one or more of the event messages is a “vote cast” event.

It is to be noted that the “terminal shutdown” message appears as the most frequent single message that is not a “vote cast” message.

It is to be noted that the most frequent quadruple of events is

0002400 PEB access failed

0002400 PEB access failed

0000706 Failed to retrieve EQC from PEB

0001635 Terminal shutdown - IPS exit

and that the normal sequence of these events seems to be that they happen in rapid succession. Sometimes the 0002400 is not repeated, and sometimes the 0002400 is repeated not twice but three times before the 0000706 and 0001635 occurs.

1. **STATISTICS AND ANOMALY REPORTS**

Finally, there are files that summarize the anomalies and exceptions that should be noted by election officials, and there are files that present histograms and frequency counts of votes cast, closing times, and such.

**beaufort\_20101102\_EXCEPTIONS.txt:**

*This is the primary file that should be used to determine what has not gone as it should in the election.* This file contains the list of a number of the exceptional events that indicate that procedure was not followed, terminals were not closed, and such.

BALLOT COUNTS:

The ballot count from the 152 and the 155 files are printed. These should be equal. If they are not, it usually means that there are terminals listed in one file but not the other. Some slight differences can also occur if the 155 file includes the provisional votes and some of those votes were ruled invalid.

TERMINALS IN THE 152 BUT NOT THE 155 FILE, OR CONVERSELY:

Exceptions are printed for the terminals that appear in one but not both the 152 and the 155 files.

MULTIPLE PEBS USED TO OPEN AND CLOSE TERMINALS:

An exception list is printed of the terminals for which the PEB that opened the terminal is not the same as the PEB that closed the terminal. These lines have terminal serial number and precinct as well as the PEB numbers for opening and for closing the terminals.

An exception list is printed of the precincts in which multiple PEBs were used to close terminals. *This appears to be the most common reason for votes to go uncounted in an election.* These lines include precinct name and number, the terminal serial numbers in use in that precinct, and the PEB serial numbers used for opening and closing terminals.

EARLY AND LATE OPENING AND CLOSING TIMES:

An exception list is printed of terminals opened early or closed late. This list would then contain all the terminals that operated all through Election Day with a grossly incorrect time and date. To avoid spurious issues with Daylight Savings Time, “early” is currently defined to be prior to 5:30am on Election Day. To take into account the need to close terminals only after 7pm, “late” is currently defined to be after 7:30pm on Election Day. These lines contain terminal number, date and time either opened or closed, and the precinct number(s) to which those terminals were assigned.

(Note that a histogram is also produced of closing times in 15-minute intervals and of votes cast on Election Day after 7pm. These histograms probably provide better information about voters still standing in line at closing time. The exception list here is intended to account for incorrect timestamps in the terminals.)

TERMINALS AND PRECINCTS WITH MISSING DATA:

There is a list of terminals whose memory cards have not been collected and a list of the precincts missing data (i.e., where the certified count is not the same as the count obtained from the 155 data), and a list of precincts missing entirely from the 155 file.

**beaufort\_20101102\_Statistics.txt:**

This file contains data and histograms that could be used by election officials to compare precincts against each other in terms of the quality of access to voting and the “smoothness” with which the election was conducted in the various precincts.

***We have presented these statistics in a somewhat crude way. It is our intent to write the data in an appropriate format and then to use a tool like Excel to create better statistical summaries. For now, the purpose is to show what data can be obtained and gather input from local election officials as to what data would be wanted and found useful. We can make it prettier, but there’s no point in spending a lot of time making output pretty unless the output is viewed as valuable, and we solicit the input of the LEOs as to what would be viewed as valuable.***

The first list of data and histogram is of votes per iVotronic terminal. We have computed the mean and standard deviation and then histogrammed the votes per terminal. We have indicated the precincts in which all terminals in the precinct have fewer votes than the mean and precincts in which all terminals have more votes than the mean. These precincts might be viewed as needing fewer or more terminals in the next elections as LEOs looked at how to allocate terminals efficiently.

For reference: If the data is distributed normally (a Bell curve), then approximately 16% of the histogrammed data will be smaller than one standard deviation less than the mean and approximately 16% of the data will be larger than one standard deviation larger than the mean. We have indicated the mean by “M” and the values “mean minus deviation” and “mean plus deviation” by “D”. The values falling either smaller than the smaller “D” mark or larger than the larger “D” mark can be viewed as the outliers.

The second histogram is of votes cast after closing time. The outliers here could indicate longer than average lines at closing time and another possible point where consideration could be given to reallocation of resources in the next elections.

The third histogram is of terminal closing times. This is in 1-minute intervals after 7:00pm on Election Day. For this data we would certainly NOT expect a Bell curve, because we have printed out a message and then skipped all machines whose closing times were way off, so there should be no “left tail” in the Bell curve. Nonetheless, the terminals closed much later than average might indicate precincts with long lines at closing or terminals that proved difficult to close.

The fourth histogram is of the actual vote event times, grouped in 15-minute intervals. In the case of the entire county, this is at present unreadable and needs to be scaled (but that’s part of making the data “pretty” and we will do that later).

This histogram of vote event times for the entire county is followed by a separate histogram of vote event times done precinct by precinct.

At the moment, we indicate this only numerically. That is interval 0 is the time period 7:00am-7:15am, interval 1 the time 7:15am-7:30am, interval 20 the time 12noon-12:15pm, and so forth.

There are still a few minor problems in the representation of this data. Most of these are that it’s not easy to determine that a terminal’s time is so far off as to exclude it from the histogram. In Beaufort County, for example, one terminal collected votes with the timestamp set to 11/29/2010, but the terminal’s time was reset to the correct time prior to closing. So the precincts with that terminal (355 and 356) have a wildly skewed histogram.

**beaufort\_20101102\_StatisticsAnomalies.txt:**

This file contains data and histograms about what we perceive to be anomalous events that bear further scrutiny. For example, there has never been a satisfactory explanation of what event log message 0001633 through 0001635, the “Terminal shutdown” messages, really mean. We have histogrammed the frequencies of these messages by terminal and by precinct in hopes of getting a better understanding of events.

Also, we have histogrammed the frequency of occurrence by terminal and by precinct of the four-event sequence mentioned above:

0002400 PEB access failed

0002400 PEB access failed

0000706 Failed to retrieve EQC from PEB

0001635 Terminal shutdown - IPS exit

We perceive this to be an instance of an inability of a PEB to communicate properly with an iVotronic terminal. There are many reasons why this might happen under normal circumstances: the PEB might be slightly off-kilter when it is slotted into the iVotronic; the PEB’s infrared signal might be angled slightly off, and/or the terminals IR receiver might be angled slightly off, so the combination of that PEB with that terminal is a bad match; or the PEB’s IR window might have become smudged.

It is also possible that these communications problems are indicative of a bad PEB or a bad terminal. Our goal in looking at these sequences of events is to try to identify failing PEBs or failing terminals so they can be sent for servicing before the next election. Unfortunately, if the initial PEB-to-terminal communication is unsuccessful, then it is unsuccessful BEFORE the terminal determines the PEB serial number, so we cannot tell which PEB in that precinct might be failing. We are trying to narrow down this by looking at similar failures with the same PEBs and the same terminals in the primary elections in hopes of being able to provide quantifiable analysis of what would be problem hardware.