HALAL Working Paper No. 7

Post-election audits: detecting electronic voting and counting machine miscounts through statistical sampling[1]

by Roberto Verzola

Abstract: This paper describes, based on existing literature on the mathematics of election audits, a Philippine-specific method of implementing post-election audits by conducting a manual recount of a sample of voting precincts. The size of the sample is determined by the total number of precincts involved in the electoral contest, the level of confidence desired, and the presumed number of precincts needed to reverse an election outcome, which in turn is determined by the total false gain in votes needed and the largest number of votes per precinct that a cheat would dare tamper with while retaining plausibility of the tampered results. While Philippine-specific, the method described here may be adapted for other countries by modifying some of the parameters such as average precinct size and the maximum plausible false gain per precinct.

It is clear from the experience of the U.S. and other countries that automating elections does not eliminate the errors and problems that have plagued manual elections. The increasing number of documented cases in various countries of errors and problems attributed to electronic voting and counting machines attest to the fallibility of these machines.

The Election Incident Reporting System (EIRS) of the Verified Voting Foundation, for instance, lists nearly 2,400 instances of machine errors during the 2004-2005 U.S. election period alone.[2] Another study found the actual error rates among electronic voting machines in current use in the U.S. far above – and therefore in violation of – the federally-mandated maximum allowable error rate of 1 in 500,000 ballot lines.[3] Typical machine errors include: failing to tally the results from some voting machines, failing to count ballots, failing to count votes for particular candidates, assigning votes for a candidate to another candidate, switching votes, vote totals that exceed the number of ballots cast or the number of registered voters, negative votes, and so on. Some reported machine errors would have changed election outcomes, had they gone undetected.[4] In a few instances, the voting machine errors probably *did* change election outcomes.

---

1 Most of the technical details on confidence-level election audits in this piece were taken from Kathy Dopp, “History of Confidence Election Auditing Development and Overview of Election Auditing Fundamentals”, Feb. 28, 2008. See http://http://electionarchive.org/ucvAnalysis/US/paper-audits/History-of-Election-Auditing-Development.pdf

2 See http://voteprotect.org

3 Brad Friedman, “Vote counting failures in Ohio and New Jersey Violate Federal Law, Accuracy Mandates”, Mar. 10, 2008, http://www.bradblog.com/?p=5884

4 Brad Friedman, “Outcomes changed in multiple Arkasas elections after widespread ES&S touch-screen voting machine failures”,May 22, 2008, http://www.bradblog.com/?p=6009

There are deeper reasons why counting and tabulation errors persist even when elections are automated:[5]

1. bugs in the software,
2. problems in the hardware,
3. environmental stresses
4. design flaws
5. human errors, and
6. malicious tampering

These factors are hardly ever absent in any automated system and are therefore a constant source of risk. Even the highest quality software will continue to contain unintended or unforeseen side effects, known as software bugs. Every software patch or update opens the door for new bugs to creep into the software. Conflicting or changing requirements tend to result in design compromises, and every compromise opens up possibilities for new errors. For instance, strict security checks may reduce ease-of-use for voters unfamiliar with the technology, forcing designers to balance the risks of relaxing security versus more user errors. Hardware calibration is often a compromise between too many false positives (wrongly attributing votes to a candidate) and too many false negatives (missing votes cast for a candidate). Even after proper calibration, environmental stresses during machine storage, deployment or under actual use can cause errors rates to deteriorate. Poor screen or ballot design can not only worsen user error rates, it can also affect errors rates per candidate differently, possibly creating a bias for one and against another. Since human intervention can only be reduced but never eliminated, voting machine problems due to designer, programmer, technician, operator, user and bureacratic errors can be expected to persist. Behind some of these problems may well be attempts at malicious tampering hiding as unintentional errors.

One way to catch voting machine errors during the counting and tabulation process itself (i.e., in real-time) is to use the time-tested methods of double-entry accounting to detect and locate routine clerical errors as well as fraudulent data masquerading as errors, as this author has proposed elsewhere.[6] Another way is to implement strict quality control of delivered machines by conducting a thorough 100% pre-election audit of every machine as well as backup, to ensure that each one meets all the performance standards set by the election authorities, in particular the minimum error rates.

This paper discusses another way to manage the risk from errors in automated elections -- to subject every election contest to a mandatory post-election audit, using statistical sampling techniques. These scientific techniques can establish to any desired level of confidence whether or not the winner declared by the automated election system is the true winner in the contest. The science and mathematics of confidence-level election audits are explained in the recent works of

[5] Roberto Verzola, "Working Paper No. 5: Automated Elections: Computers Have Made Mistakes Too", June 20, 2008.

[6] Roberto Verzola, "HALAL Working Paper No. 2: Double-entry Accounting Can Make E-Voting Systems Software Independent", May 30, 2008.

Aslam, Popa and Rivest,[7, 8] Stark,[9, 10] and Dopp.[11] The approach has been tried in Cuyahoga County of Ohio and has been officially recommended for use in California.[12]

The main difference between this type of audit and the currently more common fixed audit is that a confidence-level audit recounts a variable number of randomly selected polling precincts, depending on the winning margin and the confidence level desired, while the fixed audit recounts a fixed percentage, say 5%, of all polling precincts involved in an election contest. The latter often results in very low levels of confidence in the audit results. Occasionally, it can result in an extremely high level of confidence, which means that the sample size was unnecessarily large and the election audit unnecessarily costly.

Conducting a confidence-level election audit

The basic procedure in conducting a confidence-level election audit is explained by Dopp:

1. **The complete machine-based results must be publicly released before the audit**, to preclude subsequent alterations that can reduce the discrepancy between the machine results and the audit results. The machine-based results should be considered unofficial until confirmed by the audit.

2. **A high target level of confidence should be adopted** that can convince most doubters about the integrity of the election results. The target confidence level recommended by the American Statistical Association president to U.S. Senator Dianne Feinstein, author of S. 1487 (Ballot Integrity Act of 2007) is 99%[13] (i.e., a one-in-a-hundred chance that a questionable election outcome is not detected by the audit). Dopp suggests 95% (i.e., a one-in-twenty chance of not detecting a questionable election outcome). Targetting higher confidence levels requires sampling more precincts and therefore makes an audit more

---

7 Javed Aslam, Raluca Popa and Ronald Rivest, "On Estimating the Size and Confidence of a Statistical Audit", June 30, 2007. See http://http://www.usenix.org/events/evt07/tech/full_papers/aslam/aslam.pdf

8 Javed Aslam, Raluca Popa and Ronald Rivest, "On Auditing Elections When Precincts Have Different Sizes" (DRAFT Jan. 17, 2008). See http://http://www.mit.edu/~ralucap/AslamPopaRivest-OnAuditingElectionsWhenPrecinctsHaveDifferentSizes.pdf

9 Philip Stark, "Conservative Statistical Post-Election Audits", Nov. 24, 2007. See http://http://statistics.berkeley.edu/~stark/Preprints/conservativeElectionAudits07.pdf

10 Philip Stark, "Election audits by sampling with probability proportional to an error bound: dealing with discrepancies" (DRAFT Feb. 20, 2008). See http://http://statistics.berkeley.edu/~stark/Preprints/ppebwrwd08.pdf

11 Kathy Dopp, "Mandatory Post-Election Vote Count Audits – How to Determine Initial Sample Size for 95% PPMEB Confidence-Level Vote Count Audits (DRAFT Feb. 25, 2008). See http://electionarchive.org/ucvAnalysis/US/paper-audits/UT/VoteCountAudits.pdf

12 Post-Election Audit Standards Working Group, *Report: Evaluation of Audit Sampling Models and Options for Strengthening California's Manual Count*, July 27, 2007. See http://www.sos.ca.gov/elections/peas/final_peaswg_report.pdf

13 Mary Ellen Bock, Letter to Senator Dianne Feinstein, July 24, 2007. http://www.amstat.org/news/2007ASAElectionLettertoDFeinstein.pdf

expensive. For comparison, most scientific conclusions are based on a 95% level of confidence.

3. **The minimum number of polling precincts[14] that need to be chosen at random and manually recounted for the audit should be computed**, based on the difference in votes between the winning candidate and the closest losing candidate as well as the level of confidence desired. Dopp identifies two methods for making the estimate:[15] a uniform method, where each precinct has as much chance as getting selected as any other precinct, and another called the Probability Proportional to Margin Error Bound (PPMEB) method. For electoral systems with wide variations in precinct sizes, the latter is is more efficient (by an average of 30%) because larger precincts have a proportionately larger chance of getting selected. But it is also more complicated and harder to understand by election officials and the public. While both methods have only been recently proposed for election audits, they have already been proven in non-electoral applications. Because the Comelec will be clustering precincts in the May 2010 national elections to have around 1,000 registered voters per clustered precinct (with one machine servicing each cluster), the simpler method suits the Philippine case just fine. A detailed procedure for estimating the sample size based on the simpler method is given in the last section of this paper ("Estimating the sample size").

4. **The precincts to be audited must be selected at random and in public**. The Post-Election Audit Standards Working Group formed by California Secretary of State Debra Bowen suggests the following criteria for a truly random and transparent selection method: a) "the method produces true 'simple random samples', i.e., every subset of precincts is equally likely to be chosen"; b) "there is no possibility of human interference"; and c) "observers can see every detail of the process". This eliminates computer-based random number generators, according to the Working Group, because "generally, so-called random number generators in computers do not generate random numbers, but rather, pseudo-random numbers. True random number generators rely on unpredictable physical phenomena, such as radioactive decay. In contrast, pseudo-random numbers are predictable." [16] Instead, the Working Group supports the recommendation of Cordero, Wagner and Dill (2006), who -- after reviewing a number of practical options such as cryptography, drawing from a box, lottery-type draws, random number charts, cards, coins and dice – proposed to use 10-sided dice of different colors whose fairness have been properly ensured.[17]

5. **A public recount of the sample precincts must be conducted**. Because of the importance of this process, the recount be conducted with utmost transparency so that all major stakeholders are convinced of its integrity.

---

[14] While Steps 3 and 4 refer to precincts as the audit units, other audit units are possible depending on specific election practices. Where a precinct contains several machines, for instance, individual machines may serve as the audit units.

[15] Kathy Dopp, "History…", see above.

[16] Post-Election Audit Standards Working Group, see above, p.23.

[17] Arel Cordero, David Wagner and David Dill, The Role of Dice in Election Results – Extended Abstract, June 16, 2006. See http://www.cs.berkeley.edu/~daw/papers/dice-wote06.pdf.

6. **If the results of the recount of the sample precincts do not attain the desired confidence level, the sample size must be increased, an additional set of samples randomly selected, and the recount repeated**, until the desired confidence level is reached. The most complete description of the technical details involved in setting sample sizes and computing the confidence level attained after the recount has been provided by U.S. mathematician Kathy Dopp.[18]

7. **Once the desired confidence level is reached, the machine results can be declared official**. In the rare case that the desired confidence level is not reached despite the audit being expanded up to a full recount (for instance, if enough ballots are missing or damaged), then another election has to be called.

Estimating the sample size

Since precincts sizes in Philippine elections are generally uniform (around 200 registered voters per precinct in earlier elections and around 1,000 registered voters per clustered precinct for the May 2010 national elections), there is no need adopt a more complicated approach based on variable precinct sizes. We propose a simpler procedure for estimating the sample size (i.e., the number of precincts to be audited), using the uniform approach suggested by Dopp and modified for the Philippine context. The procedure is described below, together with a hypothetical example of the winning presidential candidate getting 600,000 more votes than the second placer.

- **Compute the difference in votes** (M) between the winning candidate (or the last among the winners) and the losing candidate (or the first among the losers). Call this difference the margin M. If this margin were the result of a miscount large enough to reverse the outcome, then the false winner must have padded his total and/or shaved votes from the true winner's total to gain at least M+1 false votes and change the election outcome. In the given example, M = 600,000

- **Estimate the maximum false gain in votes per precinct** (V) that cheats will dare to attain, while keeping the tampered results plausible. This paper suggests for the Philippine situation 25% of the total registered voters per precinct (100 out of 200 in previous elections; 500 out of 1,000 for the 2010 elections).[19] This target false gain of 500 votes may be obtained either by padding the cheat's total with 500 false votes, by shaving from the winner's total 500 votes, or by both padding and shaving to gain a total of 500 false votes. In the Philippines, a common

---

[18] Kathy Dopp, M.S. Mathematics, *Mandatory Vote Count Audit - A Legislative & Administrative Proposal*, February 2006 (last updated July, 2008). http://electionarchive.org/ucvAnalysis/US/paper-audits/legislative/VoteCountAuditBillRequest.pdf

[19] This suggestion is based on the expected Philippine clustered precinct size of 1,000 registered voters in 2010. It is not only country-specific, it is also election-specific. In previous elections, the typical precinct size in the Philippines was around 200 registered voters. In countries where precinct sizes vary over a larger range, the precinct size variation and other factors have to be taken into account.

precinct-level padding method is to add a digit to a low two-digit number; a common precinct-level shaving operation is to omit a digit from a three-digit number; a common cheating method that both pads and shaves votes is to switch the votes between the cheat and the winner, for a false gain that is twice the original winner's margin. Note how this number V was arrived at: we are presuming that to minimize the danger of detection by an audit, cheats would try to operate in as few precincts as possible. They can do this by trying to get a false gain that is as large as possible in the precincts where they will operate, but without making the manipulation so obvious the results become implausible. For the purposes of the May 2010 post-election audit, we suggest using V=500. A different number mutually agreed upon by the major stakeholders in the process may also be used.

- **Compute the minimum number of precincts returns P that must have been tampered** to change the outcome. P can be computed based on the equation $P = \lceil (M+1)/V \rceil$, where the enclosing symbols "⌈ ⌉" mean "round up to the next whole number".[20] The equation means: to gain a total of M+1 false votes while maintaining the plausibility of the tampered results, field operators must gain an average of V false votes in P precincts. In the given example, P = (600,00+1) ÷ 500 or P = 1,200.002 precints (round up to 1,201 precincts.[21] This is the minimum number of precinct results that must be tampered with, falsely gaining at least V votes per precinct, to get a total false gain of at least M+1.

- **Decide on the confidence level desired** (L) that a corrupted precinct will be detected by the audit. We propose the same 95% level of confidence as scientific conclusions, i.e., L = 0.95.

- **Finally, compute the sample size** S (the number of precincts that must be randomly selected and then manually audited). The following equation must be solved for S, given the total number of precincts N:

$$L = 1 - \quad 1 - \frac{\frac{(N-P)!}{S!*(N-P-S)!}}{\frac{N!}{S!*(N-S)!}} \quad = \quad 1 - \frac{(N-P)!*(N-S)!}{N!*(N-P-S)!} \qquad (1)$$

where S = sample size (number of precincts to be randomly selected for audit)
N = 75,471, the total number of clustered precincts in the May 2010 elections[22])

[20] Some journal articles on election audits base their formula not on the maximum false gain, but the maximum voteshift, defined as votes subtracted from the winner's total and simultaneously added to the cheat's total, yielding a false gain that is double the voteshift. This paper prefers to use the more general concept of false gain, because cheats do not necessarily use the voteshifting method and even when they pad and shave at the same time, the number of votes shaved and padded are not necessarily equal.

[21] The term "precinct" is used here synonymously with "clustered precinct", for brevity.

[22] Suarez, E.T. "Comelec has 6,739 spare machines". *Manila Bulletin*. Jan. 29, 2010. See http://www.mb.com.ph/articles/240422/6739-surplus-poll-machines.

P = minimum number of precincts needed to alter election outcome
L = level of confidence that the audit detects at least one tampered precinct result

The equation above cannot be easily solved algebraically for the sample size S. Aslam, Popa and Rivest have proven that the following conservative approximation can be used:[23]

$$S = \left\lceil [N - \frac{(P-1)}{2}] * [1 - (1-L)^{\frac{1}{P}}] \right\rceil$$

In the given example, if the winning margin was due to manipulation that was widespread enough to change the final outcome, then the false winner would have had to gain 600,001 false votes over his nearest rival.

These 600,001 falsely gained votes would have had to be distributed among many precincts, in a way that the false gain in any single tampered precinct result will still make the result appear plausible. Based on our proposed assumption that no more than 500 votes per clustered precinct will be tampered with to maintain plausibility, then the 600,001 false votes must be distributed over 600,001 ÷ 500 or 1,200.002 precincts, which we will round up to 1,201 precincts. Thus, by gaining an average of 500 false votes each in 1,201 precincts out of the 75,471 precincts involved in the contest, it would be possible to gain a total of 500 x 1,201 or 600,500 false votes, enough to reverse the election's outcome. The field operators that would implement the tampering in each of the target 1,201 precincts may reach their 500-vote quota by adding at least 500 false votes to their candidate's total, shaving 500 votes from the true winner's total, or using any of various pad-and-shave methods ("dagdag-bawas" in local parlance) such as vote-switching (called "voteshifting" in some journals). It is of course possible to gain more than 500 false votes per precinct over more than 1,201 precincts (and face greater risk of detection), but 500 votes over 1,201 precincts is the minimum needed to reverse the outcome while minimizing the risk of detection.

A post-election audit must reliably detect, to the desired level of confidence, *at least one* of these 1,201 tampered precincts among the 75,471 precincts. It is clear that the larger the sample size, the greater the chances that at least one of the 1,201 tampered precincts will be among those sampled. In fact, the audit must target a high level of confidence (say, 95%, or L=0.95) that at least one tampered precinct result will be included in the sample.

We will now use the Aslam, Popa and Rivest formula to compute the desired sample size S for the given example:

$$S = \left\lceil [N - \frac{(P-1)}{2}] * [1 - (1-L)^{\frac{1}{P}}] \right\rceil$$, where N=75,471; P=1,201; and L=0.95

[23] Aslam, Popa and Rivest, see above.

$$S = \left\lceil [75471-\frac{(1201-1)}{2}]*[1-(1-0.95)^{\frac{1}{1201}}] \right\rceil = \lceil 186.52 \rceil = 187$$

Thus, by selecting 187 precincts at random from the 75,471 precincts involved and recounting the ballots in these precincts, we can be 95% certain that we will detect at least one tampered precinct, if indeed 1,201 precincts results out of the 75,471 had been altered significantly enough (500 false votes or more) to change the election's outcome. If, in the audit, we detect no discrepancy of such extent, then we can be 95% certain that no precinct result tampering significant enough to change the election outcome had occurred. Then, the winner can be official proclaimed.

More complicated methods have also been proposed that take into account variations in precinct sizes.[24] This paper consider the Aslam, Popa and Rivest method described above sufficient for the purposes of the May 2010 national elections in the Philippines. The method needs only a scientific calculator or spreadsheet program that can compute exponents or logarithms.

Please note that the method described here assumes that only the precinct returns were tampered, but not the ballots that would be manually recounted.

Examples

Example 1: Winning presidential candidate has a 600,000-vote margin over the second placer and the total number of precincts is 75,471 (see text above for details)

M=600,000
N=75,471
V=500
L=0.95
P=(600000+1) ÷ 500 (rounded-up) = 1201
S=187

Suppose the margin of the winning presidential candidate over the second placer is 600,000 votes. If a random sample of 187 clustered precincts shows no discrepancy between the machine and manual audit counts that is at least 500 votes, then we can be 95% confident that the automated system accurately reflects the outcome of the presidential elections.

Example 2: No. 12 (winning) senatorial candidate has a 60,000-vote margin over the no. 13 (losing) candidate and the total number of precincts is 75,471

---

24 See, for instance, Javed Aslam, Raluca Popa and Ronald Rivest, *On Auditing Elections When Precincts Have Different Sizes*, January 17, 2008 (Draft). http://people.csail.mit.edu/rivest/AslamPopaRivest-OnAuditingElectionsWhenPrecinctsHaveDifferentSizes.pdf

M=60,000
N=75,471
V=500
L=0.95
P=(60000+1) ÷ 500 (rounded up) = 121
S=1,845

Suppose the margin of the no. 12 senatorial candidate over the no. 13 candidate is 60,000 votes. If a random sample of 1,845 clustered precincts shows no discrepancy between the machine and manual audit counts that is at least 500 votes, then we can be 95% confident that the automated system accurately reflects the outcome of the senatorial elections.

Example 3: Winning mayoralty candidate has 2,000-vote margin over the second placer and the total of precincts is 1,000

M=2,000
N=1,000
V=500
L=0.95
P=(2000+1) ÷ 500 (rounded up) = 5
S=450

Suppose the margin of the winning mayoralty candidate over the second placer is 2,000 votes. If a random sample of 450 clustered precincts shows no discrepancy between the machine and manual audit counts that is at least 500 votes, then we can be 95% confident that the automated system accurately reflects the outcome of the mayoralty elections.

Conclusion

Countries which decide to automate their counting of votes should consider adopting legislation that requires mandatory confidence-level election audits by election authorities, if they want to minimize the risk that the inevitable machine errors that occasionally crop up as well as fraudulent attempts to manipulate election results may change the outcome of the elections. Sample language for such legislation has been drafted by Dopp.[25] The entire audit process should be open to all stakeholders – in particular to the candidates themselves and their representatives, to election watchdogs and observers, to the media and to the general public.

A 100% pre-election audit of every machine can weed out the problemmatic ones early in the automated election process. The adoption of double-entry election tabulation will enable the real-

[25] Kathy Dopp, *Amendment Suggestions for the Senate Ballot Integrity Act of 2007 (S1487) To Increase Public Verifiability of Election Outcome Accuracy and Eliminate Unfunded Mandates*, Sept. 2007. http://electionmathematics.org/em-legislation/S1487Amendments.pdf

time audit of an election tally, whether the tally is manual or automated. Finally, the adoption of confidence-level election audits as proposed here will further reduce the risk that errors made by electronic voting or counting machines or tampered precinct results may have changed the outcome of an automated election. Another method would be required to detect tampering of the ballots themselves.

Through these and various other means of ensuring clean and honest elections, the public can become more confident in the integrity of a fundamental foundation of any democratic society.

30 May 2008
Updated June 20, 2008
Updated July 14, 2008
Updated September 12, 2008
Updated January 31, 2010

[Roberto Verzola (rverzola@gn.apc.org) is an engineer by training and is involved in efforts to reform the system of election tabulations in the Philippines. He is currently the secretary-general of Halalang Marangal. This article was first written during his six-week research fellowship at the Oxford Internet Institute (OII) of the University of Oxford in the UK on April 21 – May 30, 2008 and was revised to make its procedures compatible with the specific requirements of the May 2010 Philippine elections.]