Imperative of Independent Source Code Review for the Next Automated Elections

Pablo Manalastas
Lecturer, Ateneo & U.P. Computer Science
IT Consultant, Center for People Empowerment in Governance
Email: <pmanalastas @ gmail . com>

Abstract

We define what constitutes a proper source code review of the AES, we prescribe what conditions are necessary for such code review to be done by political parties and interested groups independently of COMELEC and the vendor of the technology, and we give reasons why it is imperative that such code review be done.

Introduction

An automated election system (AES) for Philippine elections consists of the computer hardware, the transmission infrastructure, the computer programs or software, the human participants (both voters and COMELEC employees), and all procedures that govern the conduct of elections, whether national or local. If inadequate computer and transmission hardware are used, together with insufficiently specified and poorly tested computer programs, by inexperienced system integrators and managers, then the resulting AES will be error-ridden, at best. A case in point is the AES-2010, the first computerized national elections in Philippine history, that was full of errors, both in the hardware and computer programs used, and in the manual aspects of the computerization process, as can be read in the CenPEG report[1]. We must, however, move on, and the lessons that we learned from our errors in AES-2010 should help us design a better AES-2013 and AES-2016.

The source code of a computer program is “code written by a programmer in a high-level language and readable by people but not computers. Source code must be converted to object code or machine language by a compiler before a computer can read or execute the program”[2]. Source “code review is systematic examination (often as peer review) of computer source code. It is intended to find and fix mistakes overlooked in the initial development phase, improving both the overall quality of software and the developers'/programmers' skills”[3]. Source code review is done by people. That is why they must use the human readable version of the computer programs, which is the source code, for their review. Furthermore, the most important objective of a source code review is to determine if the programs conform to their specifications - that is, to check if the programs do what they are supposed to do. In the case of the AES for Philippine elections the computer programs must be proper implementations of our election laws, and so must conform to the computerization provisions contained in RA-9369 [4] and related documents, including COMELEC's Terms of Reference (TOR) to bidders of the 2010 computerized National and local elections

Aims of Paper

The aims of this paper are (1) to define what constitutes a proper source code review of the AES, (2) to prescribe what conditions are necessary for such code review to be done by political parties and interested groups independently of COMELEC and the vendor of the technology, and (3) to give reasons why it is imperative that such code review be done.

What Constitutes Proper Code Review of the AES?

The source code review done by SysTest Labs [5] of the AES-2010 included syntax checks using automated tools to determine proper coding practices, to determine if the code conforms to the US EAC 2005 VVSG [6] coding standards. SysTest Labs found many bugs, which to date have not been resolved, including errors in database transactions, variables allocated without being freed, integer variables of insufficient width, etc. These bugs actually resulted in serious errors in the COMELEC tabulations, as can be observed from the COMELEC public access website http://electionresults.comelec.gov.ph/ [7]. In this website, one finds that of the 76,000 total number of clustered precincts, there are 21,766 clustered precincts in which 32,455 positions have no results. That is, the results for 21,766 clustered precincts show that from one to three candidate positions do not have published results. Instead, in the space for results, one reads the unusual error indicator "$contestResult" as shown in this page for clustered precinct CP271 in Antipolo City.

Now this public access website shares program code in common with the PCOS and CCS computers, the Precinct Count Optical Scan (PCOS) computers used at the precincts and the Consolidation and Canvassing Servers (CCS) used for the municipal/provincial/COMELEC/congressional canvasses, and this code sharing is well-documented in the SysTest Labs report [5]. So you can imagine what kinds of errors lurk within the PCOS and CCS programs that the public has no chance to see.

We mentioned earlier that the ultimate purpose of certification is to check that the AES hardware, software, etc are proper implementations of our election laws. Nowhere in the SysTest Labs report is there a mention of any provision of RA-9369 or the COMELEC TOR against which the AES-2010 was checked for conformity. SysTest Labs totally ignored our laws!

A proper source code review should, therefore, consider provisions of our laws and rules that are relevant to the computerization process. We enumerate here the relevant provisions, and how the source code review should treat such provisions.

1. Accuracy in Counting and Canvassing

Section 7 of RA-9369 states as a minimum system capability " x x x (b) Accuracy in recording and reading of votes as well as in the tabulation, consolidation/ canvassing, electronic transmission, and storage of results; x x x". The term "accuracy x x x in the tabulation" should be taken to mean not just the ability to read the shade marks manually put by the voter in the ovals on the ballot, but more

importantly, the ability to credit that vote mark to the candidate chosen by the voter. There are a number of steps from the recognition of a voter's shade mark on the ballot to the assignment of vote to the proper candidate, and the ballot designers and AES computer programmers could make mistakes in any of these steps. To illustrate the ease with which errors can come in, one only needs to recall the events of May 3, 2011, when during the final testing and sealing stage of the preparations [8] for the May 10, 2010 elections, it was discovered that the names and row-column positions of candidates on the ballot did not match the names and row-column positions of candidates in the configuration data of the PCOS computer, as read from CF-cards. The ballots were modified and updated many times as new candidates were included in the elections and as old candidates were disqualified, and while the ballots were being modified, the election managers forgot to write the modified data to the CF-cards for the PCOS computers. To complicate matters, this mismatch between ballot data and PCOS data occurred in a majority of the 1,600 distinct municipal/district election contests that had to be configured into 76,000 PCOS computers, which will be used by 51 million voters. To complicate matters further, each ballot was about 8.5"x24" and had two faces, the front face had 266 names of candidates [9] for national positions, and the back face had a smaller number of candidates for local provincial and municipal positions. To complicate matters even further, it is a well established fact that no government agency can deliver the corrected CF cards from Manila to the far off barangays on islands that are far from Manila in less than five days. The magnitude of the numbers is mind-blowing, and no single-day computerized election has been tried anywhere in world with such numbers.

We propose that to ensure "accuracy in the tabulation", that the 1600 ballot faces, and the 1600 PCOS configuration data be made part of the source code review, in order to check for proper match. Such configuration data would ordinarily be part of the source code, even if the data are read from CF card, and so should be part of a proper source code review. These detailed check would not be needed in a referendum in Venezuela, where the only choices are "Yes" and "No", nor would such checks be needed in a Democratic party primaries in the U.S., where the only choices are "Obama" or "Palin". But detailed checks would need to be made in the Philippines, where there is that seriously confusing choice of 266 names of national candidates, and several dozen choices of local candidates, in each of the 1,600 municipalities where 51 million voters will cast their votes.

2. Voter Verifiability of His Choices

Section 7 of RA-9369 states that an AES with minimum system capabilities should, "x x x (n) Provide the voter a system of verification to find out whether or not the machine has registered his choice; x x x". After the voter feeds his ballot to the PCOS computer, the PCOS must show the voter, on the LCD screen or on printed paper, how the PCOS interpreted his vote marks, by showing him a listing of the names of candidates that the PCOS interpreted as his votes, and then allow the voter to make changes whenever the voter does not agree with the PCOS's interpretation, and finally the PCOS must print such voter-initiated corrections on the ballot, or do such actions as the COMELEC might specify as proper corrective action. In addition, although not part of the law, the PCOS should have provision for the voter to verify, after he has finished casting his ballot, that his votes have been included in the official count, using a system like Scantegrity [10], such that by checking the COMELEC website, the voter can get positive proof that his votes have been included in the official count.

We propose that a source code review should check that these capabilities, namely voter verifiability that the computer has correctly registered his choices and voter verifiability that his votes have been included in the official count, are actually part of the source code.

3. Ensuring Integrity of Transmission of Election Data

Section 8 of RA-9369 specifies that "All electronic transmissions by and among the AES and its related components shall utilize secure communication channels x x x to ensure authentication and integrity of transmissions". We are a country consisting of several thousand islands, and so a computerized elections with electronic transmission will depend on a working communication network infrastructure. But we do not have a communication network that covers 100% of the country 100% of the time. So we must devise a hardware-software-manual solution that will work 100% of the time. We propose a three-stage solution as follows: (1) transmission via TCP/TLS via the Internet will be tried first. (2) If this does not work, then after a short period of wait time, say W minutes, the same data will be cut up into SMS sized fragments to be transmitted via SMS/GSM. (3) If there is not even a GSM signal, then the digitally-signed election data on portable back-up media (SD/MMC/CF cards or USBsticks) will be transported by courier (COMELEC employees) to the canvassing centers. At the receiving end, no matter how the data arrived, either by means of (1), (2), or (3), the election data will be treated the same way, will be included as part of the canvass using the same computerized procedures and will be transmitted further on the canvassing chain and saved to back up servers as if they all arrived in the same manner. This way, there will be no inconsistency in the COMELEC databases between election data transmitted electronically, and election data carried by human couriers.

A proper source code review should check that the computer source code for both the PCOS and CCS computers include provisions to implement solutions (1), (2), and (3), because such solution is a third-world solution that will work for a country that does not have 100% connectivity like the Philippines.

4. Provision for Proper Digital Signing of Election Documents

Section 19 of RA-9369 specifies that "The election returns (ER) transmitted electronically and digitally signed shall be considered as official election results and shall be used as the basis for the canvassing of votes and the proclamation of a candidate". This provision is further specified in the COMELEC Request for Proposal (or Terms of Reference – TOR) to bidders of AES-2010 in Section 4 on "Counting, Consolidation and Generation of ER" which states that "x x x 4.5 The BEI shall digitally sign and encrypt the internal copy of the ER x x x". While the law RA-9369 specifies that the ER has to be digitally signed, the COMELEC TOR clarifies that the digital signing has to be done by people, by the members of the Board of Election Inspectors (BEI), which is the intention of the Omnibus Election Code (Batas Pambansa-881).

Current standards for digital signing employ public key cryptography (PKC). Under PKC, a person, Juan Cruz, who wishes to digitally sign electronic documents (computer files) must generate a pair of crytographic keys (e.g. using the program OpenSSL), a private key which Juan must keep secret and not divulge to anyone, and a public key which he submits, together with documentation on his identity, to a Certificate Authority (CA) like Verisign, who will create (for a fee) a digital certificate (a computer file) stating that the given public key really belongs to Juan Cruz. Then the CA and Juan Cruz can broadcast to the world Juan's digital certificate containing his public key. If Juan Cruz is a member of the BEI of a voting precinct in Antipolo City, it is the responsibility of the CA and Juan Cruz to supply a copy of Juan's digital certificate to COMELEC, to the Board of Canvassers in the Province of Rizal where Antipolo City is located, to all political parties, and to the public, who will use his public key to check the authenticity of the precinct ER that Juan will sign.

A proper source code review must prove that BEI member Juan Cruz can do digital signing of the precinct ER without divulging to the PCOS machine his private key. Current technology for doing this

requires Juan Cruz to install his private key and a signing program in a Processor Smart Card (PSC) [11], which is just a smart card with a little computer that can digitally sign using Juan's private key. In order to digitally sign the precinct ER, the PCOS computer computes the SHA* hash value of the precinct ER, then outputs this hash value to Juan's PSC. The signing program in Juan's PSC then encrypts the hash value using Juan's private key, and outputs the encrypted hash value back to the PCOS computer. Juan is thus able to digitally sign the precinct ER without his private key ever leaving his PSC.

A similar digital signing procedure of the Statement of Votes (SOV) and Certificate of Canvass (COC) must be installed in the CCS computers that will be used for canvassing.

A proper source code review must reveal that the method of digital signing described here is part of the computer programs of the PCOS and CCS, as required by the digital signing provision of law.

5. Correctness of Transmission and Canvassing Plan

A proper source code review must reveal that each PCOS computer has a unique identification (uuid if a Unix machine) and unique place of usage as a voting machine, namely the precinct and barangay and municipality and district and province where it will be used. Also, code review must reveal that the CCS server to which the PCOS will transmit its precinct ER is the correct one.

Furthermore, source code review must reveal that each CCS computer has a unique identification (uuid if a Unix machine) and unique place of usage as a canvassing machine, namely the municipality or district or province or others, where it will be used. Also code review must reveal that the CCS is extracting that portion of the received election documents (ER or SOV or COC) that it needs to do its own canvass, and that the CCS server to which it will transmit its SOV and COC is the correct one. For example, the CCS server in Congress, although it receives transmissions of all precinct ER and all municipal district and provincial SOVs and COCs, must only use the portion of the provincial COCs that it needs to do the canvassing of presidential and vice-presidential votes, and just file away all other received documents, ignoring them for its own canvassing.

Confirming the correctness of this transmission and canvassing plan as part of source code review will prevent the occurrence of such glaring errors as the reporting of a total of 256 million voters at the Joint Congressional canvass, when the actual total is only 51 million [12].

6. Integrity of Voting and Canvassing Programs

Section 11 of RA-9369 specifies as a function of the Technical Evaluation Committee (TEC) the certification of the AES based on documented results, that include, “x x x 5. A certification that the source code reviewed is one and the same as that used by the equipment; x x x”. We want to go a step further and require that the source code reviewed is one and the same as that used by the equipment, and that this fact is voter verifiable anytime on election day. We suggest inclusion in the program running on the PCOS a button-activated function that computes the SHA* hash value of the code-segment (non-modifiable read-only portion) of the running program, and displays this hash value to the voter, so that he can compare this hash value with the one created during the trusted build which must be pasted on the PCOS at a convenient location.

A proper source code review should verify that such hash computation is part of the source code of both PCOS and CCS computers.

7. Disambiguation of Test Runs from Actual Election Runs

Programs must be written on both PCOS and CCS computers that label transmitted data as data of a test run, or as actual election data, so that results of testing are not included in the official canvass. This inclusion of test run data from more than 200 precincts in the official canvass happened during the May 10, 2010 elections [13], and Congress approved the inclusion, because the "transmitted results are the official basis of canvass and declaration of winners".

Source code review must reveal that such disambiguating programs are present in both PCOS and CCS computers.

8. Voter-Friendly Rejection of Ballots

The program running on the PCOS computer must be voter friendly enough to explain to the voter why his ballot was rejected. Every case of rejection must be accompanied by a courteous, friendly, and clear explanation why the ballot was rejected. Explanations like "The PCOS could not read the bar code on the ballot because of possible extraneous marks in the bar code area", "This ballot is for another precinct because of wrong ballot serial number", "Complete failure of PCOS to read the ballot, possibly due to PCOS malfunction", are so much better for the voter's confidence in his own intelligence than a silent rejection of his ballot by the PCOS, without any explanation.

A proper source code review should reveal that rejection of ballots is done in a voter-friendly way.

What Conditions are Necessary for Independent Code Review?

There are very few conditions necessary for political parties and interested groups to do their own source code review independently of COMELEC. First is the grant of the right to such groups to study the source code, under an environment of complete freedom, free from artificial restrictions imposed by COMELEC or the technology vendor selected by COMELEC. But this grant of right to study the source code is already in our laws. Section 12 of RA-9369 states, "x x x Once an AES technology is selected for implementation, the Commission shall promptly make the source code of that technology available and open to any interested political party or groups which may conduct their own review thereof". All that is needed is for COMELEC to honor this right of the people.

Second is the freedom to publish the results of such source code review, so that the people may know how the AES will count and canvass the votes.

Finally, there must be freedom from suits that third parties might file against the source code reviewers. To prevent suits of this kind, the COMELEC must require technology vendors who want their AES technology to be used for Philippine AES, to offer their software products under an "open source license", such as the GNU General Public License, or BSD license. Vendors who offer their products to be used for Philippine AES must derive their profit from the sale of hardware, and not from licensing of software, since the Filipino public who will be reviewing the source code will be contributors to the improvement of vendor software.

Why It Is Imperative that Independent Code Review be Done

Source code review done by political parties and interested groups independently of COMELEC is necessary in the exercise of our democracy.

First, freedom of information is in our Constitution, and the freedom to study the source code of the AES is in Section 12 of RA-9369. These provisions in our laws are acknowledgement that computer programs for administering elections are fast at counting and canvassing, at the sacrifice of transparency of the process. People do not understand how computers count and canvass our votes, because they do not witness the count. And so, our laws compensate by letting the computer programmers among our people study the source code of the computer programs so that they can tell the rest of the people that the election computer programs are correctly counting and canvassing our votes, according to our election laws, and they can rest assured in this knowledge that they are not being cheated. And there is sufficient number of computer programmers among our people that they are in great demand by technology companies, both here and abroad.

Second, COMELEC needs the help of the IT community of the Philippines, even if it is too proud to admit this fact. COMELEC does not have enough IT-competent people in its staff to do a proper job of computerizing our elections, and for this reason, the IT community must lend a helping hand.

Third, elections work only because the people trust the system being used. The only sure way that the people will trust the system being used is to allow them to study the system, in an environment of freedom from dictation and restriction from COMELEC.

Conclusions

Proper source code review must not only check for correct programming practices, but must check for conformity of the AES computer programs to our election laws. For political parties and interested groups to do source code review independently of COMELEC, it must be guaranteed enough freedom to study the source code in an "open source" environment, without the danger of harassment of lawsuits from third parties. It is imperative that source code review be done independently of COMELEC, because the people can not trust a system that they do not understand.

References

[1] The CenPEG Report on the Philippine National and Local Elections of May 10, 2010 is available from http://www.cenpeg.org/The%20CenPEG%20Report/The_CenPEG_Report.html

[2] Definition of source code is given here: http://www.thefreedictionary.com/source+code

[3] Definition of source code review is given here: http://en.wikipedia.org/wiki/Code_review

[4] Republic Act 9369 is available from the site: http://www.chanrobles.com/republicactno9369.html

[5] The SysTest Labs report entitled, "Certification Test Report for Source Code Review, Readiness and Security Testing: Philippine AES Voting System", dated February 9, 2010, describes the source

code review done by SysTest Labs in “4 Source Code Review” on pages 8-27.

[6] This is the United States Election Assistance Commission (EAC) 2005 Voluntary Voting System Guidelines (VVSG). The guidelines are downloadable from the webpage, http://www.eac.gov/testing_and_certification/2005_vvsg.aspx

[7] The website http://electionresults.comelec.gov.ph/ has been taken down at present. However, a number of NGOs have made mirror copies of that web site, and will publish these mirrors for the public.

[8] This ABS-CBN news article, and subsequent articles, describes the error caused by the mismatch between row- column positions of candidates on the ballot, and row-column positions of candidates contained in the CF card configuration data: http://www.abs-cbnnews.com/nation/05/03/10/errors-force-comelec-reset-pcos-testing

[9] The ballot used for Philippine Election 2010 for non-ARMM regions contained 266 candidates' names on the front face. See http://www.comelec.gov.ph/downloadables/2010official%20ballot %20asof- 0208/national_nonARMM.pdf

[10] The following article describes the Scantegrity II system, in which the random numbers are revealed to the voter using a special ink for marking his choices on the ballot: http://www.usenix.org/event/evt08/tech/full_papers/chaum/chaum_html/ScantegrityII.html

[11] Java processor SmartCards are described in this article: http://www.javaworld.com/jw-12-1997/jw-12- javadev.html

[12] The error of 256 mllion voters is reported in http://www.abs-cbnnews.com/nation/05/24/10/house-server- shows-256-million-filipino-voters. Another error of 153.9 million registered voters is reported in http://votereportph.org/blog/total-registered-voters- printed-national-canvass-report-153902003

[13] The more than 200 cases of Final Testing and Sealing (FTS) results being included in the canvasses are listed here, http://pmana.multiply.com/journal/item/177/Congressional_Canvass_in_Quandary_Jun_03_10, and those FTS results reached all the way up to the national Congressional canvass