Post-Election Audits in the Philippines

Carsten Schürmann

Center for Information Security and Trust
IT University of Copenhagen
carsten@itu.dk

Abstract. How do you observe the unobservable? The election technology in use in the Philippines are optical ballot scanners called Vote Counting Machines (VCMs) that scan, count, and transmit election results at the close of polls back to the national tallying center. Post-election audits called Random Manual Audits (RMAs) are required by law to take place prior to the result becoming final. In this paper, we explore the idea of replacing RMAs by Risk-Limiting Audits (RLAs) that are efficient, have a high chance of correcting an incorrect election outcome by the means of a recount, and can therefore strengthen public confidence in the election.

1 Introduction

How do you observe the unobservable? Election technologies handle voter information, ballots, and results in digital form. To observe the processing of a ballot requires the observer to follow the flow of electrons in a system that comprises billions of transistors and millions of lines of code. This is clearly impossible!

The Philippines uses election technologies for vote casting, vote counting, and also results transmission. A voter hand-marks a ballot paper by filling-in ovals with a black pen before putting it into a so-called *privacy sleeve* and proceeding to the vote counting machine (VCM) of the clustered precinct. A cluster precinct consists of several precincts and serves up to 800 voters. A VCM is an optical ballot scanner that stores and tabulates the results. Differently from other ballot scanners, the VCM produces also a VVPAT (voter verifiable paper audit trail) that is a printout of the interpretation of the ballot by the VCM. The voter is invited to check the VVPAT and deposits it then in a special VVPAT box. The Cast Vote Record, i.e. the interpretation of each ballot cast on the VCM in digital form, and other information such as configuration files and log files are stored on two SD cards, a main card and a backup card. After the poll closes, the VCM is used to transmit the results to various servers and produces multiple printouts of the election record (ER), i.e. national and local returns, and audit logs.

To assess *election integrity*, we should remind ourselves, that election integrity cannot be evaluated by inspecting the election technology alone. An optical ballot scanner, such as a VCM, may have software defects hidden deeply inside the system, or it may misbehave, because a malicious actor might have gained access

to the system prior to the election, for example through exploiting vulnerabilities, supply-chain, or other cyberattacks, and manipulated its software. What can be observed, however, is the evidence that is produced for and by the VCM: hand-marked paper ballots and the VVPAT. Both are voter-verified, the ballot papers are hand-marked by the voter, clearly representing the voter's intention, the VVPAT can be checked by the voter after the ballot has been scanned to ensure the that the scan was successful. To check this evidence, the Philippines Statistics Authority (PSA) conducts a Random Manual Audit (RMA) after every election as required by law.

In contrast, driven by the use of various election technologies in the U.S., post-election audits have become in recent years a major area of research, which includes the theory and statistics of post-election audits [3], as well as techniques to make the usable [4]. One technique that stands out are risk-limiting audits (RLAs) that are designed to confirm election results by drawing and inspecting random samples of ballots.

In this paper, we explore if the RMA could be implemented by a risk-limiting audit (RLA). In contrast to an RMA, which requires one ballot box per congressional district to be chosen randomly and recounted manually, an RLA will draw a sample of ballots at random based on the desired level of confidence. An RLA is one of the few if not the only auditing technique that will automatically correct an incorrect election result with high probability by triggering a full hand-count of all ballots if necessary. We consider two flavors of RLAs, ballot-polling audits and ballot-comparison audits.

Hypothesis: If the post-election audit for the Philippines general election would require an RLA instead of RMA, the audit of the election outcome would be more (1) expressive, (2) autocorrecting, and (3) more efficient, if we consider previous elections.

The paper is organized as follows. In Section 2, we summarize the state of post-election auditing in the Philippines and describe the legal framework and implementation of RMAs. We then introduce briefly RLAs in Section 3, before we consider and evaluate the impact of RLAs in the previous Philippines election in 2016 and 2022 in Section 4. Next, we assess results and conclude with Section 5.

2 Random Manual Audit

In this section, we describe the current situation in the Philippines, including the legal framework, the technique that is used to select at random a polling station in a congressional district, and finally the process of conducting the audit. With the introduction of VCMs into the voting process, *clustered precincts* were defined that comprise several "traditional" precincts, which means that up to 800 voters can use one and the same VCM.

2.1 The Legal Framework

The election law authorizing the use of an automated election system (AES) for the Philippines general election can be found in *Republic Act No. 9369*, approved

23 January 2007, which is an act amending Republic Act No. 8436, entitled "an act authorizing the commission on elections to use an automated election system in the May 11, 1998 national or local elections and in subsequent national and local electoral exercises, to encourage transparency, credibility, fairness and accuracy of elections, amending for the purpose Batas Pambansa Blg. 881, as amended, Republic Act No. 7166 and other related election laws, providing funds therefor and for other purposes.

Besides providing the legal justification for the use of technology, the law also governs the use of post-election audits, which are called Random Manual Audits (RMAs) in the Philippines. The relevant paragraph reads as follows:

> SEC 29. Random Manual Audit. - Where the AES is used, there shall be a random manual audit in one precinct per congressional district randomly chosen by the Commission in each province and city. Any difference between the automated and manual count will result in the determination of root cause and initiate a manual count for those precincts affected by the computer or procedural error.

There are 243 congressional districts in the Philippines.

To implement the provision of the law, the Commission on Elections (COMELEC) promulgated Resolution 10774 on March 23, 2022 amending Resolution 10738 promulgated on Dec. 9, 2021, entitled "In the Matter of the General Instructions for the Conduct of the Random Manual Audit (RMA) for the [May 9, 2022] Automated Synchronized National and Local Elections and Subsequent Elections Thereafter."

Resolution 10774 requires that "the actual number of precincts to be selected in a legislative district shall be determined by proportional allocation, that is, based on the number of clustered precincts a legislative district has in proportion to that of all the other legislative districts in the country."

The law states that the audit may take up to 45 days.

2.2 Election Results for 2022

We focus our attention to the presidential (see Table 1) and vice-presidential race (see Table 2). Results for the other 9 races can be found online.

2.3 Drawing a Random Sample

The random sample of clustered precincts to be audited was chosen by software that was developed by the Philippines Statistical Authority (PSA) and reviewed by third parties.[1] As a result 757 clustered precincts were selected [2] in the presence of media and observers, out of which 746 ballot boxes were eventually audited and 27 were subjected to further verification, because the content

[1] See `https://www.manilatimes.net/2022/06/15/opinion/columns/random-manual-audit/1847437`

[2] See `https://comelec.gov.ph/?r=2022NLE/RandomManualAudit2022`

MARCOS, Ferdinand Jr. Romualdezos	31,629,783	58.77%
ROBREDO, Maria Leonor Gerona	15,035,773	27.94%
PACQUIAO, Emmanuel Dapidran	3,663,113	6.81%
DOMAGOSO, Francisco Moreno	1,933,909	3.59%
LACSON, Panfilo Morena	892,375	1.66%
MANGONDATO, Faisal Montay	301,629	0.56%
ABELLA, Ernesto Corpus	114,627	0.21%
DE GUZMAN, Leodegario Quitain	93,027	0.17%
GONZALES, Norberto Borja	90,656	0.17%
MONTEMAYOR, Jose Jr. Cabrera	60,592	0.11%
Total Votes	53,815,484	

Table 1. Presidential Race Philippines 2022

DUTERTE, Sara Zimmerman	32,208,417	61.53%
PANGILINAN, Francis Nepomuceno	9,329,207	17.82%
SOTTO, Vicente III Castelo	8,251,267	15.76%
ONG, Willie Tan	1,878,531	3.59%
ATIENZA, Jose Jr. Livioko	270,381	0.52%
LOPEZ, Emmanuel Sto Domingo	159,670	0.31%
BELLO, Walden Flores	100,827	0.19%
SERAPIO, Carlos Gelacio	90,989	0.17%
DAVID, Rizalito Yap	56,711	0.11%
Total Votes	52,346,000	

Table 2. Vice-Presidential Race Philippines 2022

of the ballot was damaged or ERs were missing. Although the software was carefully reviewed, some stakeholder groups publicly distrusted that the selection of clustered precincts was random.[3]

2.4 Conducting RMAs

An audit comprises a manual tally of all 11 contests on the ballot and judgments about what is a valid mark and what is not. Considering the voter turnout of about 83.07%, the expected number of ballots to be audited is around 503,071. The logistical effort for arranging an audit of this magnitude are immense. Ballot boxes must be transported to the Manila where the audit is executed, and since the ballot contains several races, a sort and count approach does not work.

[3] See `https://www.change.org/p/the-truth-petition-manifesto-exhorts-the-comelec-to-open-750-randomly-selected-ballot-boxes-for-manual-count-and-audit-of-sd-cards-sign-and-share-this-petition-now-click-here-bit-ly-truthpetitionph`

Instead, the information of the ballot is carefully recorded by other means, and an accuracy score is computed.

For the 2022 election, the accuracy score was determined to be 99.95928%. COMELEC reported[4] that out of 757, a total of 746 ballot boxes were audited. Some ballot boxes were no longer subjected to audit, while 27 are still subject to further verification of the Technological Evaluation Committee for the following reasons: mislabeled ballot boxes, with wet/torn ballots, and no printed and online election returns. The root cause of the discrepancies, we suspect, was due to a difference in interpretation of manual vs. automatic interpretation of the hand-marked ovals on the ballots.

3 Risk-Limiting Audit

A risk-limiting audit (RLA) [3] refers to a family of post-election auditing techniques that confirms a correct or corrects an incorrect election result with high probability, which is given by the risk-limit. It is a technique that reduces the trust in the correctness of the election result to the trust in the security of the evidence, usually hand-marked paper ballots, machine-marked paper ballots, or VVPATs.

The workings of the RLA and the reason why it works is best explained by an analogy[5]. If we were to determine if a large pot of soup is too salty, nobody would expect us to drink the entire pot: it is sufficient to stir the soup well and then take a spoonful. In the analogy, the soup represents all ballots, the spoon a sample, the "saltiness" the margin between winner and runner-up, and the tasting the verification. In a risk-limiting audit, the risk-limit defines how certain we want to be that the election result is correct, the size of the spoon is determined by statistics, and the stirring of the soup by picking a truly random sample. If the sample is not random, the result of the RLA will hold no truth.

If the RLA cannot confirm the election result, it triggers a full hand recount, and this recount will deliver the correct result. The RLA brings efficiency and, recognizing the challenges of stakeholder trust in smaller sample sizes, integrity to post-electoral audits. Different social choice functions require different techniques, for example, standard ballot-polling or ballot-comparison audits apply to first-past-the-post voting schemes, such as the one used in the Philippines, but there are also others that apply to the d'Hondt voting rule [5] and Single Transferable Vote (STV) systems [2].

3.1 Ballot-polling Audit

For a first-past-the-post system, the auditor conducting a ballot-polling audit selects a truly random sample of ballots and counts them. When the votes provide sufficient evidence that the election result is correct, the audit stops, otherwise

[4] See `https://www.pna.gov.ph/articles/1177078`

[5] Credit to Prof. Philip Stark, personal communication.

```
function draw_sample(totalvotes, samplesize, entropy):
  for i = 1 to samplesize:
      x = entropy ^ "," ^ i
      y = hash(x)
      z = lookup(y mod totalvotes)
      print(z)
  end
```

Fig. 1. Drawing a truly random sample

the sample size is increased until a full hand count of the ballot papers is triggered. Ballot-polling audits are not the most efficient audits, but they will work for any first-past-the-post election. A more efficient RLA is a ballot-comparison audit, which we discuss next.

3.2 Ballot-comparison Audit

Following [3], ballot-comparison audits confirm an election outcome by comparing hand counts to voting system counts for clusters of ballots. Comparison audits can be thought of as having two phases: (i) Check whether the reported subtotals for every cluster of ballots sum to the contest totals for every candidate. If they do not, the reported results are inconsistent; the audit cannot proceed. (ii) Spot-check the voting system subtotals against hand counts for randomly selected clusters, to assess whether the subtotals are sufficiently accurate to determine who won. If not, the audit has a large chance of requiring a full hand count.

3.3 Drawing a Random Sample

Whether ballot-polling or ballot-comparison audits, the math behind RLAs will determine the initial sample size to be drawn based on the risk-limit given. We present a technique in Figure 1 for drawing this sample, which is truly random and publicly verifiable: To draw the sample, entropy is collected, which is often done using ten-sided dice in conjunction with a cryptographically secure hash-function `hash`. The technique works well when ballots are identifiable. In the Philippines each ballot is uniquely identifiable by a barcode, which contains information such as the polling place identifier and a ballot serial number. Next, each ballot identifier is transcribed using the ballot manifest into the relevant precinct and serial number information (using the function `lookup`) and subsequently printed (using the function `print`), as outlined in the code below. Based on this information ballots should then be physically retrieved and checked.

The use of a cryptographically secure hash function guarantees that the algorithm is verifiable: If the manually generated entropy is known, anyone with

Fig. 2. Entropy collection

Legislative District/ City/ Municipality/ Province/ Region	Polling Place/ Address/ Barangay	Clustered Precincts	Ballot identifier
Maguindanao - first City of Cotabato Maguindanao Barmm	Lugay - Lugay Central School Kibatang St. Lugay - Lugay Bagua I Bagua	0155A, 0158A, 0161A, 0162A	295
Sulu - first Patikul Sulu Barmm	Kaumpang Elementary School Bangkal, Patikul Igasan	0060A, 0061A, 0062A, 0063A	137
...	...	...	...

Fig. 3. A sample list of ballots to be audited

a computer and limited programming skills can compute and verify that the set of audited ballots is correct.

For example, for the 2022 presidential race, where $53,815,484$ ballots were cast and a sample size of 49, we first collect entropy as displayed in Figure 2. The sample can be computed using `draw_sample(53815484, 49, "674987539")`. For illustration purposes, Figure 3 depicts a hypothetical output. Note that the right most column denotes the ballot to be checked in the clustered precinct identified in the third column. Different entropy generates different lists.

Note, that the method `draw_sample` could be used as an alternative to the way how precincts are selected in an RMA (see Section 2.3) that chooses a truly random sample of precincts among the $412,874$ used during the Philippines election. To use the method proposed here, generate new `entropy` and run `draw_sample(412874, 757, entropy)` with an appropriate `lookup` function that turns numeric precinct identifiers into precinct names. This method has several advantages over the method used in RMAs, the most important of

which being that the verification of the software or the software itself does not need to be trusted.

3.4 Executing the RLA

Executing an RLA is straightforward.

In the case of a ballot-polling audit, ballot after ballot is drawn following the sample set computed in the previous section. Once all ballots were retrieved, and it was determined that they statistically support the election result, the audit stops, otherwise, the RLA will increase the sample set to be audited.

In the case of a comparison-ballot audit, the ballot under audit is drawn and then compared against its digital interpretation in the cast vote record, which is originally stored on the SD cards of each VCM and later integrated into a comprehensive database.

Drawing a ballot implies that the auditors will need physical access to the hand-marked paper ballots or, alternatively, the VVPATs.

3.5 Correcting an Erroneous Outcome with an RLA

In the case that the election outcome is not confirmed the RLA algorithm may either increase the size of the sample or call immediately for a full hand-count. A full hand-count is easier and more efficient to organize and execute than to locate and verify each and every ballot individually. Recall, that the sample size depends on the margin between winner and runner-up and on the risk-limit. The greater the risk-limit, the smaller the sample size. A full hand-recount will determine the correct result and help identify the root cause for any discrepancy that might have occurred.

4 Evaluation

The conditions in the Philippines are well-suited for conducting either a ballot-polling or even a ballot-comparison audit against the cast vote record: Paper evidence is secured, voters appear to have confidence in the security of the paper trail, and there is already an understanding that audits are useful and should be conducted. The authorities could either audit the hand-marked paper ballots or the VVPATs. In general, we would recommend using the hand-marked paper ballots, because they most closely represent the intent of the voter, which renders the value of VVPATs redundant for the purpose of election integrity. We recognize of course that the VVPATs presented an efficient tool for voters to strengthen their confidence into that the VCMs interpreted their respective voting choices correctly.

Given a specified risk-limit, the efficiency with which an RLA could audit an election is determined by the margin between the winner and the runner-up. The wider the margin, the less evidence is needed to check the result, the smaller the sample of ballots to be audited. In contrast, the smaller the margin, the more

DUTERTE, Rodrigo	16,601,997	38.99%
ROXAS,Mar	9,978,175	23.43%
POE, Grace	9,100,991	21.37%
BINAY, Jejomar	5,416,140	12.72%
SANTIAGO, Miriam Defensor	1,455,532	3.42%
SENERES, Roy Sr. V.	25,779	0.06%
Total Votes	42,578,614	

Table 3. Presidential Race Philippines 2016

ballots need to be audited. This can also lead to the paradoxical case that for a given risk-limit the number of ballots that have to be audited exceed the number of ballots cast in the context.

For a better demonstration of these issues for the two different RLA methods discussed earlier, we present here also the election results for the 2016 Philippines elections, noting the margin for the 2016 election is 263,473 ballots (because of the vice presidential race), whereas the margin for the 2022 election is two orders of magnitudes larger, i.e. 16,594,010 ballots. The official results of the presidential and vice-presidential races are depicted in Table 3 and Table 4, respectively.

ROBREDO, Maria Leonor Gerona	14,418,817	35.11%
MARCOS, Ferdinand Jr. Romualdezos	14,155,344	34.47%
CAYETANO, Alan Peter	5,903,379	14.38%
ESCUDERO, Francis	4,931,962	12.01%
TRILLANES, Antonio	868,501	2.11%
HONASAN, Gregorio	788,881	1.92%
Total Votes	41,066,884	

Table 4. Vice Presidential Race Philippines 2016

We should expect that the sample size for 2016 is much larger than for 2022. Using the election auditing tools that Prof. Philip Stark offers on his webpage[6], we compute the different ballot sizes for a ballot-polling and ballot-comparison at different risk-limits. The results are summarized in Table 5 and Table 6, respectively. For 2022, if we compare the sample sizes of either RLA with the expected 503,071 ballots audited in the current elections, we observe that the RLAs are orders of magnitude more efficient. A ballot comparison audit, for example, requires only 49 ballots to audit while guaranteeing that an incorrect election outcome will be identified with a likelihood of 99.9%.

[6] See `https://www.stat.berkeley.edu/~stark/Vote/auditTools.htm#`

If we focus our attention to 2016, we note that the margin between winner and runner-up is very small. Consequently, we expect the sample size for either audit to be much larger than for 2022, and indeed it is. A ballot-polling audit still requires a substantial sample to be drawn, even if the risk limit is set to 10%. The comparison ballot audit, however, can yield 99.9% certainty that the outcome is correct, by only considering a sample of 2586 ballots.

Risk limit	2016	2022
10%	80,872	44
5%	105,169	57
2%	137,287	73
1%	161,583	85
0.1%	242,294	126

Table 5. Ballot-polling RLA. Sample sizes

Risk limit	2016	2022
10%	862	18
5%	1183	22
2%	1491	29
1%	1724	33
0.1%	2586	49

Table 6. Ballot-comparison RLA. Sample sizes

When comparing RMAs and RLAs, one key difference is that the random sample required to be inspected in an RLA may originate from any ballot box. Note, when doing a ballot comparison RLA, we do not have to recount the entire ballot box, all we have to do is to locate *the* ballot as specified by the RLA and compare it to its digital representation in the cast vote record. This means that in the worst case, with a risk-limit of 5%, in 2016, we would have to open 1,183 ballot boxes.

5 Conclusion

The requirement stipulated by the legal framework to audit election results that were produced using election technologies, such as VCM's, is a testimony for the Philippines to strive for transparent and verifiable elections. The Random Manual Audit (RMA) required by law is well-intended, but its efficiency and statistical relevance most likely could be further strengthened by considering ideas present in modern post-election technologies, such as risk-limiting audits.

To learn about the challenges of RLAs in the context of Philippines elections, the COMELEC could consult with the Philippines Statistical Authority (PSA) and derive a plan to run a RLA pilot in parallel the RMA for the next election. The logistics behind such an audit are challenging, especially when sample sizes are big.

In summary, an RLA works as follows: For a given risk limit, an RLA will, if the margin is suitably large, be an extremely efficient method to implement post-election auditing. If the margin is small, however, an RLA might even require a full hand count of all ballots, which may be justified if the desired risk-limit is small. If COMELEC ever considers implementing RLAs, the main question to be answered, is what is a suitable risk-limit and what kind of RLA should be used. Because of the availability of the cast vote record, a ballot comparison audit is possible, and should therefore be preferred.

As described, the sample sizes can be very small when conducting a risk-limiting audit, so small in fact, that voters may no longer trust the audit. Although the statistics is sound and the mathematics behind risk-limiting audits has been stress tested by several mathematicians, small sizes can give raise to distrust [1]. It is therefore advisable to evaluate to what extent voters trust the security of the paper trail and if they accept sample sizes that are as small as the ones described here.

References

1. Asmita Dalela, Oksana Kulyk, and Carsten Schürmann. Voter perceptions of trust in risk-limiting audits, 2021.
2. Floyd Everest, Michelle Blom, Philip B. Stark, Peter J. Stuckey, Vanessa Teague, and Damjan Vukcevic. Auditing ranked voting elections with dirichlet-tree models: First steps, 2022.
3. Mark Lindeman and Philip B. Stark. A gentle introduction to risk-limiting audits. *IEEE Security & Privacy*, 10(5):42–49, 2012.
4. Mayuri Sridhar and Ronald L. Rivest. k-cut: A simple approximately-uniform method for sampling ballots in post-election audits. In *Financial Cryptography Workshops*, 2019.
5. Philip B. Stark and Vanessa Teague. Verifiable european elections: Risk-limiting audits for D'Hondt and its relatives. *USENIX Journal of Election Technology and Systems (JETS)*, 1(3):18–39, December 2014.