CITIZENS' OVERSIGHT PROJECTS
CitizensOversight.org

White Paper: Election Audit Strategy

Validated and secured ballot images provides the lowest cost with acceptable risk while promoting public participation in the election process

Ray Lutz
Citizens Oversight, Inc.
Sept 10, 2018

(VERSION 0.5 -- DRAFT FOR COMMENT)

Table of Contents

1 Introduction

Auditing elections is an essential element in thwarting attacks on our democracy, providing evidence that the results are complete, honest and trustworthy.

Not all states implement audits of election results and the auditing approach and implementation varies considerably. The goal of this paper is to provide a review of recently promoted auditing approaches and provide common approach to strike a balance between efficiency and cost effectiveness. This proposed approach is a "hybrid" of several approaches that are available.

2 Background

Because a wide-range and variety of readers of this document is anticipated, this background is provided as context. Those who already are familiar with this background can skip to Section 3 "Audits."

2.1 First Generation Voting Machines: Punch Cards or Mark-Sense

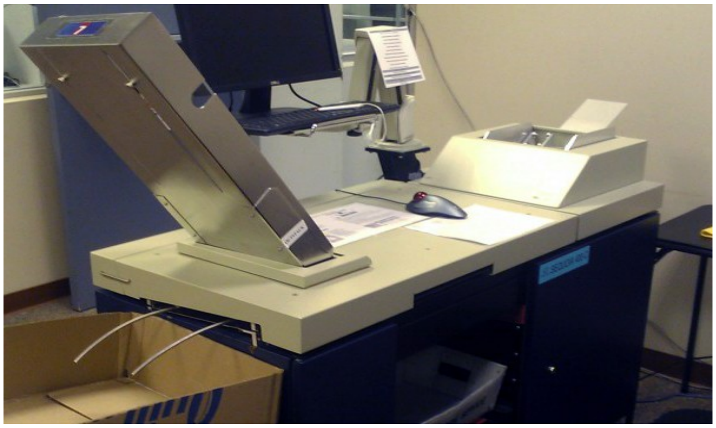


Figure 1: Sequoia (Dominion) Optech 400C

After the 2000 presidential election, the Help America Vote Act (HAVA) was passed, providing the impetus to “upgrade” many of the voting machines used nationwide. Many election districts upgraded from punch cards to ballots marked with a marking device like a pen or pencil. These technologies are roughly equivalent from a scanning standpoint. Ballots are passed through a scanner that can detect individual holes in the cards or bubbles filled in on the ballot. These “mark-sense” scanners use low-resolution scanning technology where a small number of photo-detectors (perhaps 2-20) are positioned linearly over the locations where holes or marks are expected in the ballot, which is moved across the photodetector array. A typical case is shown in Figure 1.

These devices are very successful in correctly and quickly interpreting voter intent on ballots marked or punched correctly, but lead to errors when the ballots are not marked or punched exactly right. True punch-cards are the easiest for simple optical scan machines to scan correctly, but have many drawbacks in terms of ease of use and may include incomplete punches.

When using this type of equipment, it is generally necessary to review the ballots in a quality-control step and pull out any ballots prior to scanning that will likely not scan correctly, such as any with extraneous marks, torn or ripped cards, etc. Of course, no matter what the scanning technology, when using paper ballots, any machine will occasionally jam and may tear or wrinkle a ballot so it can't be reliably scanned. Torn or ripped ballots are routinely duplicated to a new ballot manually by election staff. Records should be maintained to allow verification that this process was completed correctly. For example, Riverside and San Bernardino Counties in CA consistently "duplicate" any marginal ballot that will likely not scan correctly to create a new scannable ballot, keeping the old ballot for any later audit review. Other jurisdictions, such as San Diego and Santa Barbara, have used white-out tape to cover any extraneous marks, claiming that it is possible to always review what was done. However, we also find that there are generally no written procedures for this step, no logs maintained, no reports produced, and no third party routinely reviews whether this was done correctly. Getting access to the ballots is nearly impossible.

2.2 Second Generation: DRE Machines with no audit trail

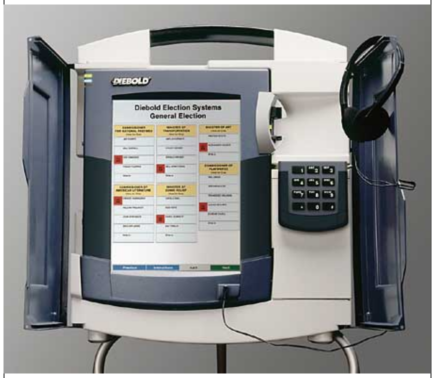


Figure 2: Diebold Accuvote TSX

Also in response to the HAVA, "Direct Recording Electronic" (DRE) voting machines were introduced, also known as "touch-screen machines." (Figure 2.) These machines do have the benefit that they can provide feedback so voters are aware of any over and under-votes, and are particularly attractive to assist disabled (such as blind) voters, and provide any of many ballot styles in any number of languages in a cost-effective manner. This is very attractive to election officials when compared with pre-printed paper ballots, where many ballots must be produced based on estimates of turnout, to ensure that each location has sufficient ballots of each language, party, and appropriate ballot style (including the appropriate races). Many ballot types are frequently required because voters in different locations can vote on different races, and the overlap of the jurisdictions can result in dozens if not hundreds of ballot types for that reason alone. Multiply this by the fact that primaries may have a separate ballot for each party and multiple languages, the number of ballot styles will make your head spin. The order of candidates listed on ballots rotates so all have the first position part of the time.

HAVA requires that at least one DRE machine be available in precincts to support the needs of disabled voters. In San Diego, although these machines exist, the number of people who use them is a very small percentage, with no one using them at all in many precincts, as voters prefer to use paper ballots. Interestingly, as of this writing, San Diego currently transcribes each person's votes from these machines onto paper ballots and scans them using a central scanner rather than accepting the votes directly from the machine.

DRE voting machines seemed at first to be a dream come true. But, they have fatal security flaws. Any voting machine that records votes directly to electronic media, like a memory card, can be hacked (or designed) so votes are recorded differently from what is displayed to the voter. It is not possible to determine later if votes were appropriately recorded. The votes from such machines could be manipulated inside the voting machine itself, prior to recording on digital media. The voter could be told one thing and then the vote recorded for the other ballot option. And if the machine failed or the digital media was lost or destroyed, all the votes in that machine could be lost. Votes in the machine could be pre-set by unscrupulous workers to favor one candidate by initializing with say +50 votes for candidate A and -50 votes for candidate B. The total number of votes is zero, but at the end of the day, the margin would be 100 votes larger in favor of Candidate A than would otherwise be the case. The main problem is that it is not feasible to go back and check the result with this type of machine with any sort of audit.

2.3 Adding VVPAT to DRE Machines

Many states that utilized these machines passed state law to require the use of a "Voter Verifiable Paper Audit Trail" (VVPAT). These are implemented typically as an add-on accessory to existing DRE machines consisting of a secured box with a window with a printer inside, which prints the voter's selections on a roll of paper. The voter -- if they spend the time -- can roll the paper back by the window so they can inspect their vote and confirm that it is correctly recorded on the roll of paper.

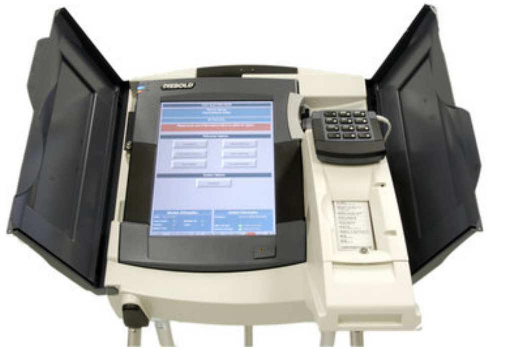


Figure 3: Accuvote TSX with VVPAT

This can help under two conditions: 1) if the voter actually reviews their vote on the paper roll and 2) if the paper roll is checked later as well. Some jurisdictions, like California, require that 100% of the paper rolls are audited and the computer result compared with the VVPAT roll. Auditing the paper rolls is logistically very difficult, and frequently, the VVPAT printers jam or run out of paper making it impossible to check the digitally recorded result.

Unless the digital version is compared with the paper trail, it is still possible for the software to manipulate the internal result.

Because of these vulnerabilities, election theorists defined the concept of "Software Independence"[1] as the need to record the votes without the use of software at all, so the actual intent of the voter could be reviewed in an audit or recount. Essentially, this means the use of durable paper ballots where the voter records their ballot selections directly.

2.4 Vote-by-Mail voting becomes popular

Meanwhile, Vote-by-Mail (VBM) voting, sometimes known as permanent absentee voting, became much more popular in areas where it was an option, especially Western states. Washington and Oregon, now utilize all-VBM elections. Election officials prefer VBM voting because the appropriate ballot can be provided to each voter, meaning that far fewer ballots need be printed of each type and language. Voters tend to like it because it is convenient. You can study the ballot alone or with your family, and even discuss the issues as you actually mark your ballot, making it somewhat like a take-home test

[1] https://people.csail.mit.edu/rivest/RivestWack-OnTheNotionOfSoftwareIndependenceInVotingSystems.pdf

rather than a pop quiz. Further, voters who have heard about the vulnerabilities of voting machines know their vote is recorded on a durable paper ballot which can be later audited, if they only have the option of voting on such machines when voting in person.

In California, additional tracking on election official's websites will allow the voter to confirm that their VBM ballot was received at the election office. In recent years, California law was amended to require that VBM ballots be accepted if they are postmarked on election day and received until up to three days later. Also, if the signature was omitted from the ballot, the registrar must attempt to contact the voter who then has up to eight days to provide their signature.

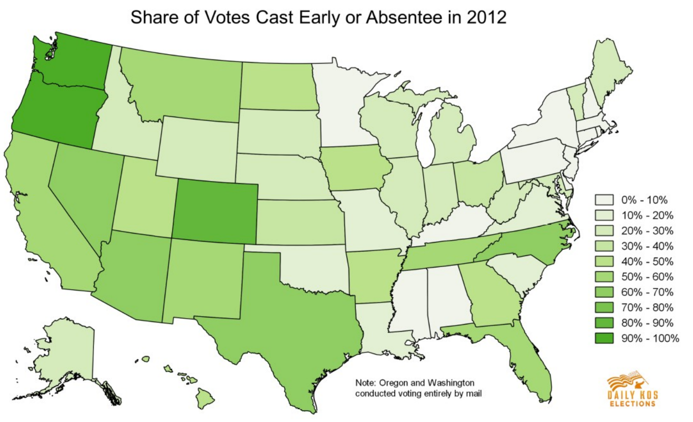


Figure 4: Prevalence of Vote-by-Mail voting

With that said, VBM has some weaknesses. First, it may allow household members or friends to strong-arm other voters. Second, people worry that ballots will be competed by friends or family members for invalids or deceased family members. Also, recently passed California law allows ballots to be picked up by non-postal and delivery organizations, like political parties, and brought to the election office. This may invite unscrupulous behavior by those who want to rig the system. But there is a limit to such fraud because only registered voters can submit a ballot and election systems will reject duplicate VBM ballots for the same voter, if that should accidentally happen.

One other implication of VBM voting is the fact that there is a time-consuming signature comparison step which is required to validate that the voter is the person they say they are and their ballot is not being completed fraudulently. Because of the popularity of VBM voting in California (more than 60% of the electorate use VBM), the state allows election officials to start processing VBM ballots ten days before election day, with the result of the tabulation sequestered in an inaccessible database so officials (and voters) will not know what the initial tally is until after the polls close. Just the fact that election workers have VBM ballots for an extended time and can reject them with matching signatures or allow ballots to be accepted with obviously differing signatures does introduce additional concerns. We find tracking these ballots is difficult for public observers. Also, election management systems frequently do not produce reports needed for effective audits.

The first totals provided at the closing of polls on election night are from the “Early VBM ballots” which are received and processed prior to election day. The tabulation is updated later that night or early the next morning with the totals from the in-person polling-place ballots. These are the “Polls ballots.” Any VBM ballots not processed by election night are processed in the days that follow and may take several weeks before they are all processed. For reference, we call these the “Later VBM ballots”.

Finally, there are provisional ballots. These occur largely due to two circumstances. First, voters who went to the wrong polling place so that poll workers cannot confirm if the voter has already voted at their home polling place. Second, VBM voters who do not have a ballot to surrender, and they may have inadvertently voted by mail and their ballot has not yet been processed. These two factors can be minimized if electronic “poll books” are used to record whether the voter has voted or not rather than paper systems.

Starting in 2018 elections, same-day registration will be offered in California, and these will also result in many provisional ballots. Registrations completed at the polling place will likely also cause a provisional ballot to be used, so the registration application can be reviewed prior to accepting the ballot. Other states are starting to offer same-day-registration to thwart rigging by purging voters or other registration scams.

2.5 Third Generation: Hybrid DRE and Paper Ballot Printer

Some voting systems are a hybrid of the DREs and paper ballot approach. The front end is like a traditional DRE, providing the benefits of error control, the support of disabled and blind voters, and the ability to provide the ballot in any language and ballot type appropriate for any voter, no matter where they may be voting. But unlike DREs, it does not rely on an internal tabulation. Instead, these systems print a durable paper ballot either exactly like a traditional mark-sense fill-in-the-bubble ballot, or one that just lists the voter's selections (rather than all the options for each ballot race.) The voter can inspect this ballot to insure it matches their intent, and then put it in the ballot box. Because the ballot is printed with predictable marking, there are no misinterpretations of voter intent during scanning. The ballot will be correctly read every time, even if a traditional mark-sense scanner is used. Some systems propose the use of a barcode (like a QR-code) to encode their vote and make it easier for the machine to extract it. But since voters are not able to know what the QR-code says without a QR code reader, it adds yet another check to be done in the process, and thus extracting the vote from the printed names is superior.

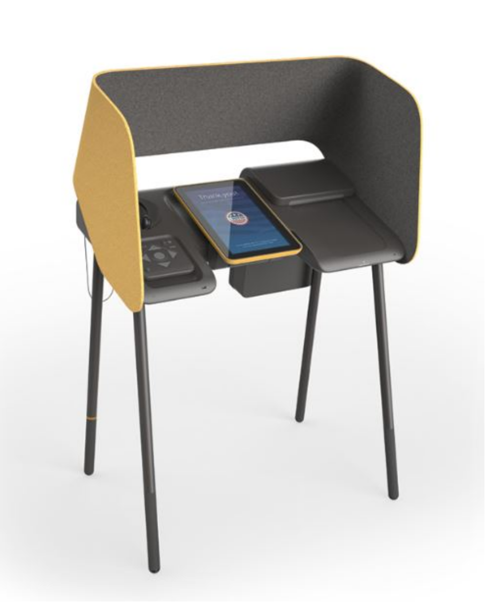


Figure 5: Los Angles County's new VSAP voting machine

This hybrid system does not affect how VBM ballots are processed. VBM voters have exactly the ballot they need based on their language and ballot type requirements because that is the ballot sent to them, and they continue to fill them out at home using fill-in-the bubble approach. The really good news for Los Angeles is they are moving away from the 60-year old "mark-a-vote" system that used tiny ballots that only had numbers on them, and the voter had to find the number of the candidate and fill it in on the ballot, yet another source of unending error and conflict.

2.6 Ballot Image Scanner -- Fourth Generation

Fourth-generation systems utilize a high-resolution image of the entire ballot rather than just specific spots on the ballot, as was the case with the mark-sense generation of election machines. For clarity, we will call these “ballot images” rather than scans[2].

We will note at this juncture that all Ballot Image scanners operate in a two step process: First, creating an image, and then, analyzing it to extract the voted selections. This is in contrast with the mark-sense type scanner that does not create an image but directly extracts the voted selections. This two step process may be processed with no perceptible delay, so it may appear to occur at the same time, or the two processes may be separated in time and location.

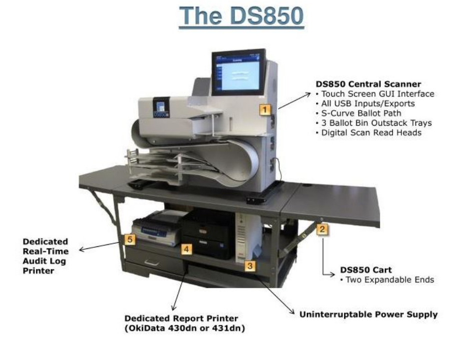

The DS850
DS850 Central Scanner
• Touch Screen GUI Interface
• All USB Inputs/Exports
• S-Curve Ballot Path
• 3 Ballot Bin Outstack Trays
• Digital Scan Read Heads
1
2
3
4
5
Dedicated Real-Time Audit Log Printer
DS850 Cart
• Two Expandable Ends
Dedicated Report Printer (OkiData 430dn or 431dn)
Uninterruptable Power Supply

Figure 6: ES&S DS850 Central Scanner

The primary benefit of using a full-ballot image is that it provides much more flexibility in terms of how marks on the ballot are to be interpreted. Mark-sense scanners only check a specific location on the ballot to see if light is reflected or absorbed, i.e. whether a hole exists, bubble is filled in, or in some cases whether an arrow is completed. Voters might not fill in the bubbles according to best practices. Printed ovals might encourage some voters to check mark or “X” the bubble or in some cases, voters complete the circle, leaving the interior blank. In these cases, mark-sense equipment might not correctly capture voter intent.

Figure 7: ES&S DS200 Precinct Image Scanner

States may have elaborate “uniform vote counting standards” regarding how voter intent is to be interpreted. California's Uniform Vote Counting Standard[3], is an example. It defines that if a voter darkens one bubble, then crosses it out, and darkens the other one instead, then this is to be interpreted as a vote for the second option. Older mark-sense equipment could never interpret this correctly on their own. They just interpret this as an over-vote and throw it out, and then require human review of overvotes to determine how it should be interpreted, perhaps only if there is a legal challenge to the results.

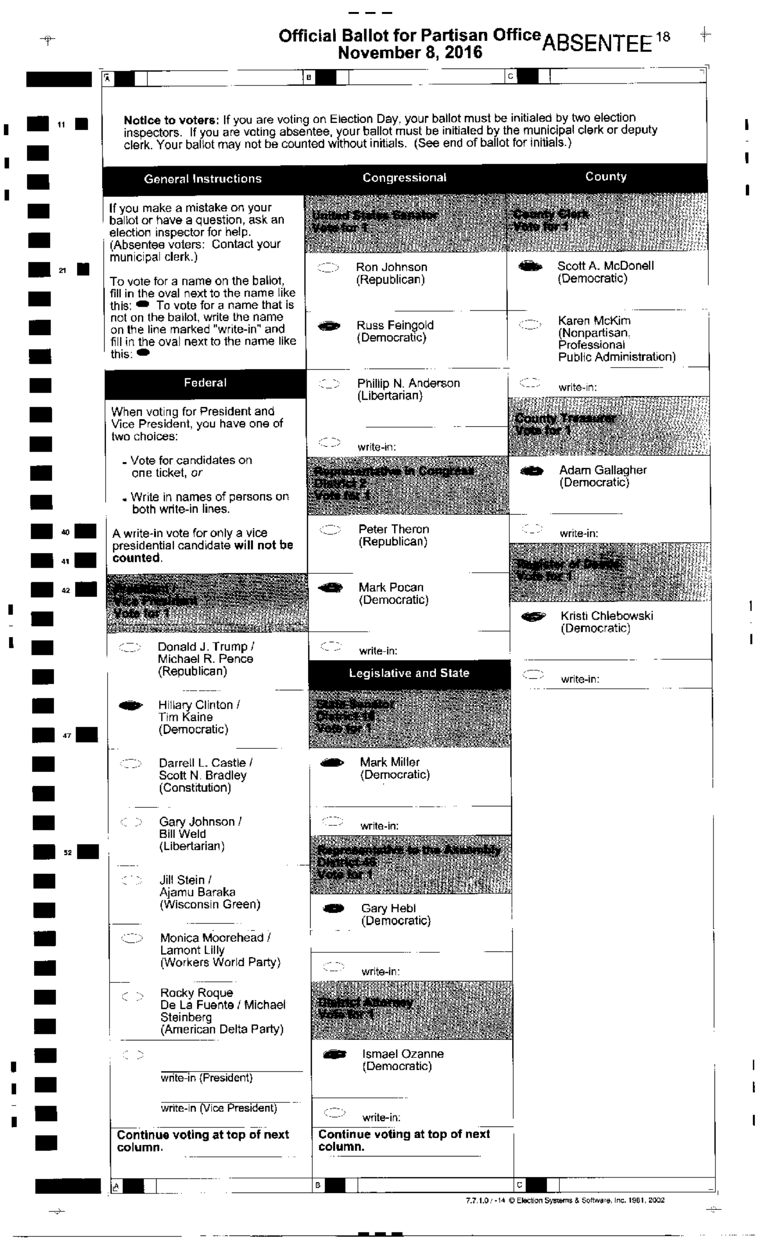
The second major opportunity provided by the use of ballot images is improved auditability. With paper ballots without ballot images, it is necessary to review

[2] We must also note that the election world has also used the term “ballot image” to mean the digital selections on a DRE machine rather than an actual visual image. We suggest that this usage be discontinued and instead call that the "Cast Vote Record."

[3] http://www.sos.ca.gov/elections/uniform-vote-counting-standards/ -- Uniform Vote Counting Standards, CA Secy of State.

the actual physical ballots to do any sort of an audit. That's vastly better than basic DRE machines that offer no means for any audits at all, but reviewing physical ballots makes audits difficult just because there is no way to automate it or spread it over many people. Physical ballots require a robust chain of custody, and we find it may take court action to allow the public to review the ballots. No matter how they are counted, one tiny mark on ballots can cause an overvote and invalidate the vote on any ballot.

Many small jurisdictions across the country still use Hand Counted Paper Ballots (HCPB) and avoid the threat of machine manipulation within a voting machine. But HCPB is not without its own set of issues. It requires a robust chain of custody to insure that ballots are not added, removed, swapped out, or modified. It is very time consuming, more difficult than anyone thinks at first, and generally subject to an error rate of about 1% to 2%. If ballots are complex with many races and ballot propositions, it is too much to ask poll workers to continue to work to count those ballots on election night, and it is very hard to provide oversight of many (thousands) of polling places.



Official Ballot for Partisan Office ABSENTEE 18
November 8, 2016

A B C

Notice to voters: If you are voting on Election Day, your ballot must be initialed by two election inspectors. If you are voting absentee, your ballot must be initialed by the municipal clerk or deputy clerk. Your ballot may not be counted without initials. (See end of ballot for initials.)

General Instructions

If you make a mistake on your ballot or have a question, ask an election inspector for help. (Absentee voters: Contact your municipal clerk.)

To vote for a name on the ballot, fill in the oval next to the name like this: ● To vote for a name that is not on the ballot, write the name on the line marked "write-in" and fill in the oval next to the name like this: ●

Federal

When voting for President and Vice President, you have one of two choices:

- Vote for candidates on one ticket, *or*
- Write in names of persons on both write-in lines.

A write-in vote for only a vice presidential candidate **will not be counted**.

President / Vice President
Vote for 1

○ Donald J. Trump / Michael R. Pence (Republican)

● Hillary Clinton / Tim Kaine (Democratic)

○ Darrell L. Castle / Scott N. Bradley (Constitution)

○ Gary Johnson / Bill Weld (Libertarian)

○ Jill Stein / Ajamu Baraka (Wisconsin Green)

○ Monica Moorehead / Lamont Lilly (Workers World Party)

○ Rocky Roque De La Fuente / Michael Steinberg (American Delta Party)

○ write-in (President)

write-in (Vice President)

Continue voting at top of next column.

Congressional

United States Senator
Vote for 1

○ Ron Johnson (Republican)

● Russ Feingold (Democratic)

○ Phillip N. Anderson (Libertarian)

○ write-in:

Representative in Congress
District 2
Vote for 1

○ Peter Theron (Republican)

● Mark Pocan (Democratic)

○ write-in:

Legislative and State

State Senator
District 16
Vote for 1

● Mark Miller (Democratic)

○ write-in:

Representative to the Assembly
District 46
Vote for 1

● Gary Hebl (Democratic)

○ write-in:

District Attorney
Vote for 1

● Ismael Ozanne (Democratic)

○ write-in:

Continue voting at top of next column.

County

County Clerk
Vote for 1

● Scott A. McDonell (Democratic)

○ Karen McKim (Nonpartisan, Professional Public Administration)

○ write-in:

County Treasurer
Vote for 1

● Adam Gallagher (Democratic)

○ write-in:

Register of Deeds
Vote for 1

● Kristi Chlebowski (Democratic)

○ write-in:

A B C

7.7.1.0 / -14 © Election Systems & Software, Inc. 1981, 2002

Figure 8: Ballot Image Sample (half size) From Dane County WI, 11,2016

It is important to mention that there are two flavors of ballot image scanning equipment and methodology. One approach is to scan and create ballot images and concurrently extract the vote from the ballots as they are processed. The equipment shown in Figure 6: ES&S DS850 Central Scanner and Figure 7: ES&S DS200 Precinct Image Scanner and equipment from other vendors such as Hart InterCivic process ballots in this manner.

The second approach decouples this process into two distinct steps. First, scan the ballots and create image files. Second, process the images to extract the vote. If done this way, Commercial Off-the-Shelf (COTS) scanners can be used. These scanners have not been explicitly designed for the election application and thus are used for many document scanning applications. This equipment is typically far less expensive and is more difficult to hack because images are created without any knowledge of the voting application (as opposed to a machine that is aware of the election application and thus has all the information needed to modify the results). Thus, this approach has slightly higher level of robustness in our view, but everything will have weaknesses and vulnerabilities.

Once the images are created and secured according to the procedure described in this document, changing the result of the election is essentially impossible without detection.

The other benefit of this approach is that the vote extraction process can be delegated to an external service which can operated using highly parallel and redundant processing, resulting in a result that can be checked and cross checked among several such services.

3 Audits

Regardless of the approach used for voting, audits of the election results are necessary to insure that the result has not been modified either intentionally or unintentionally.

A primary constraint of election audits is that they must be easily observed and confirmed by the public. Even better are independent audits by multiple third parties, if they can access secure election records. To perform any sort of audit on the election results, there must be voter verified and secured records that can be reviewed.

Ballot images can serve as a robust security feature to back up paper ballots, as paper ballots can be subjected to fire, flood, intentional destruction, and alteration. It is difficult if not impossible to detect whether a ballot has been modified on its own. If the ballot image set can be located which corresponds with the ballot, it can serve as a check on the informational content of the paper ballot, and vice versa.

There is indeed a chance that ballot images will be altered so they do not reflect the ballots, so a check on the integrity of the ballots is necessary by comparing with physical ballots. This process and how it compares and fits with other auditing methods is the central point of this paper.

3.1 Cast-Vote Record (CVR)

The Cast-Vote Record (CVR) is the record of the extracted votes of an individual ballot. The set of all CVRs in the election can be summed to provide the complete tabulation. A CVR can be related to a specific paper ballot[4] and also to the ballot image file(s). To make it easy to compare CVR, Image, and paper ballot, it is mandatory that a unique identifier is provided on the ballot which can be seen in the image and which can be extracted and included in the CVR. This unique identifier should not be a sequence number that may be utilized to link the ballot to the voter and thereby violate voter anonymity.

[4] We will refer to the paper ballot in the singular, even if it includes multiple pages. Also, ballot image may likely include several images perhaps in several files.

An additional report is the "Manifest" which allows a particular ballot to be easily located within the physical ballot set.

3.2 Audit Types

3.2.1 Rescan with check of undervotes and overvotes

Some jurisdictions perform an automated rescan of races that are very close and perform a review of the undervotes and overvotes that are detected. So if a voter voted for fewer (undervote) or more options (overvote) than were allowed, then those ballots are separated out by the scanner and they are manually reviewed to see if there is any possibility that these could be interpreted differently. In Florida, this process is called the "manual recount of overvotes and undervotes,"[5] and occurs automatically in any race where the margin is less than or equal to 0.025% ("one-quarter of 1 percent"). This is not a true audit because the automated machine could be set up to make the same changes to the vote in the recount as in the prior tabulation. However, sometimes the ballots are reviewed manually to find the overvotes and undervotes rather than using the machine. They thus assume that the machine count was accurate for the rest of the votes, and this really does not provide a check on the operation of the machine.

3.2.2 Batch Comparison Audit

The most popular type of audit is the batch-comparison audit, where each batch of ballots selected for the audit is hand-tallied, and compared with the computer report for that batch. The "1% Manual Tally" in California[6] and the "Voting System Audit" in Florida[7] are batch-comparison audits. The benefits of this type of audit is that it is fairly simple to implement, it provides a check for some types of fairly extensive hacks of the election, and it provides a direct check that the equipment is working correctly. Ideally:

(a) all batches in the election should be subjected to the random selection process, with about the same likelihood of being selected,

(b) the batches of ballots should be kept in the secure chain of custody and not handled prior to the manual tally,

(c) the full set of computer results, broken down by batch, should be frozen prior to the random selection and manual tally, and

[5] Florida Statute §102.166 "Manual recounts of overvotes and undervotes." -- http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&URL=0100-0199/0102/0102ContentsIndex.html

[6] California Election Code, §15360 -- https://codes.findlaw.com/ca/elections-code/elec-sect-15360.html

[7] Florida Statute §101.591 "Voting System Audit" -- http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=0100-0199/0101/Sections/0101.591.html

(d) the number of batches should be greater than the number of scanning machines so as to check them all. In some counties (San Diego, Santa Barbara) the number of scanners is far greater (10x) than the number of batches sampled so there is no way to cover all the potential vulnerabilities.

This audit as implemented in CA and FL is not strong enough to guarantee detection of even some very extensive hacks to any significant degree from a statistical standpoint. The probability of catching a hack depends on how many precincts are affected by the hack.

Assuming a race with two candidates, that race would need to have a close margin for any hack to be feasible, with nearly 50% of the votes already cast for the desired candidate. The votes that could be affected in any precinct would be the other (just over) 50%. To affect more than about 10% (i.e. 20% of the remaining margin) or it would be considered too obvious to attempt. In any case, the hack must be spread over a number of precincts.

If we consider the case of a district with only 100 precincts (so the result can be extrapolated to larger districts), one precinct will be chosen for the audit, and each precinct is 1% of the total number of votes. As mentioned, it is our opinion that at most 10% of the total ballots could be affected in any one precinct by a hack that would not be obvious. If those votes are flipped in the precinct (add one vote to the desired candidate while subtracting one from the other), it is like moving the election by 0.2% for each precinct included in the hack. Considering a total move of 5%, then 25 precincts (5/0.2) would be required to implement the hack. Choosing one precinct out of the 100 means there is a 25% chance that one of the 25 hacked precincts would be chosen, and the hack therefore detected. This is far lower than the 95% confidence level normally considered a goal in other sampling scenarios. To reach that level of confidence would require that more than 10% of the precincts included instead of 1%. Some of these assumptions could change but no matter what, sampling 1% of the precincts provides a low confidence level to detect this type of hack.

However, this system is not truly random, as the behavior of the hacker is dependent on whether he thinks he might be caught. If the audit is conducted correctly, it would present a risk that the hack would be detected. With this fact known by the fraudster, he would likely not attempt the hack at all. For this reason, an audit process may be successful in thwarting attacks even if the risk of being caught is not as significant as would otherwise be necessary if the process was purely stochastic in nature.

The reality is that the 1% manual tally is poorly implemented by many counties, most often by a) not subjecting all batches to the possibility of random selection, b) not freezing the computer report in advance, and c) not honestly reporting the results. For example, batches with discrepancies are sometimes simply rescanned and compared with the new computer report (and the old report with the discrepancies ignored and covered up). This is dishonest reporting, but it is also like not freezing the computer report.

As defined, this type of audit (as implemented in CA) does not escalate even if discrepancies are found, and the actions required of the election officials if any discrepancies are found is not clearly defined. As a result, these audits tend to be mostly theater rather than substance. Intense public scrutiny can help keep the officials honest. These drawbacks could be rectified by implementing procedures which would require escalation to include more batches to be tallied if the result was very close and/or if a significant number of discrepancies were detected.

There are other variations of these audits. In Florida, each county chooses only one race, and performs a batch-comparison audit on 1% of the precincts related to that race. Unfortunately, they do not perform any audit at all if they also perform a "manual recount of overvotes and undervotes" of any race in the county. This audit process is very weak and is essentially nonexistent if an "automatic recount" occurs.

3.2.3 Risk-Limiting Audits

Any audit process that is designed to escalate and require additional checking as the margin grows tighter, most particularly if there is any evidence of any discrepancies, can be called a Risk-Limiting Audit. The term should be applied only if the election records are further reviewed if there is any indication that the result may be incorrect.

This term has been used recently primarily to refer to ballot sampled audits that pull out individual ballots at random and compare the result based on the sample with the tabulated result. However, the term can be applied to any properly escalating audit procedure.

These sampling audits either (1) randomly sample the physical ballots and derive a result that can be used to validate the election results, or (2) compare each ballot with the CVR. The sampling size is determined so that there is a small chance that the election results will in fact be wrong to the extent that the wrong winner is declared.

3.2.4 Ballot-Sampled "Polling" RLA

The simplest ballot-sampled RLA is called the "Polling" RLA (PRLA). It requires the least amount of additional information in terms of any matching CVRs or having a detailed manifest. If you were presented with a dozen pallets of ballots in the corner of a warehouse and you were told to validate the election, a ballot-sampled PRLA would be a good choice. To perform this audit, pull individual ballots randomly from the ballots that comprise the election, tabulate that sample, and compare with the final result.

The tighter the race is, the more samples are required. The number of samples is inversely related to the margin in a 1/x relation.

The following graphs were generated based on data derived using the "Tools for Ballot-Polling Risk-Limiting Election Audits" online calculator[8]. There it says that "This page implements some tools to conduct "ballot-polling" risk-limiting audits as described in "A Gentle Introduction to Risk-Limiting Audits (AGI)", by Lindeman and Stark[9], and "BRAVO: Ballot-polling Risk-limiting Audits to Verify Outcomes," by Lindeman, Stark, and Yates.[10]

[8] https://stat.berkeley.edu/~stark/Java/Html/ballotPollTools.htm
[9] https://www.stat.berkeley.edu/~stark/Preprints/gentle12.pdf
[10] https://usenix.org/system/files/conference/evtwote12/evtwote12-final27.pdf

Here, we examine only a simple case of two candidates, and using a 5% risk limit. The term "candidates" is used, although this could just as easily be any sort of ballot measure with two options. A 5% risk limit is chosen because that is the limit proposed in recent legislation in CA.

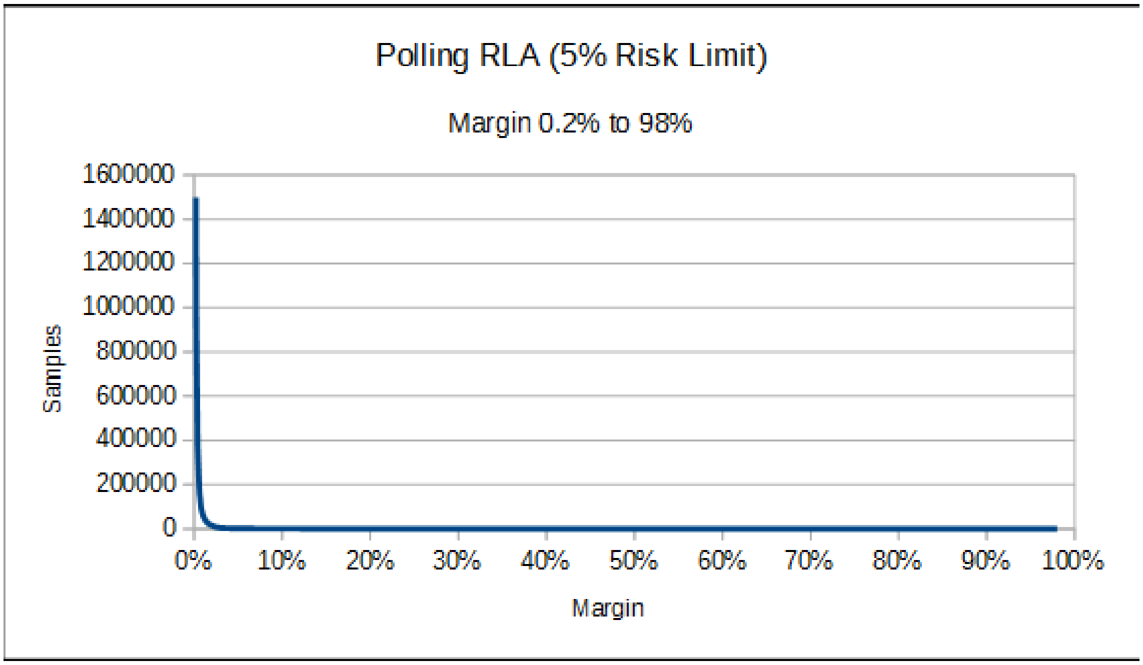


Figure 9: Samples required for a ballot-sampled polling RLA with 5% risk limit

It turns out that if you assume fully random sampling, the number of samples required is only related to the margin, no matter how many votes are included in the tabulation. We can see here that the curve follows the typical 1/x relation where it is infinite at 0%, and at 100%, it will be nearly zero (6 samples.) The number of samples becomes very large when the races get close. A direct result of this fact is that any statistical approach will be most appropriate for very large districts and will wind up being a poor match for the majority of districts which will be too small for the approaches to be implemented cost effectively. (This paper will cover this point in more detail later.)

As the margin gets smaller, once the number of samples exceeds the total number of votes, then some other auditing method is called for. Most of the literature suggest that the next step should be a "full hand-count." But in fact, it will be more economical to use some other auditing method long before that, because pulling paper ballots at random from the full set of ballots is much more expensive (in terms of time and effort -- probably on the order of 14 to 25 minutes per ballot to pull and include in the random result) than a full hand tally done assuming all ballots will be reviewed from the get-go (where each ballot can be processed in about 4 minutes -- but this figure depends greatly on the number of races on each ballot), or other more automated procedures based on digital ballot images which can be pulled and analyzed by computer (taking on the order of far less than a second to process a ballot.)

Let's look more closely at the knee of the curve, between the margins of 2% to 20%. We can see that the inflection point is at about 6% margin -- not a very close election, where just under 2,000 random ballots would need to be pulled. At smaller margins, the RLA starts to require a very quickly increasing sample size. For margins below 2%, you may as well not even start the ballot-sampled RLA, as you need to sample 14,000 ballots (or much more!) A 1% margin requires that 60,000 ballots be sampled. Due to the cost of pulling ballots, this approach does not compete with other approaches.

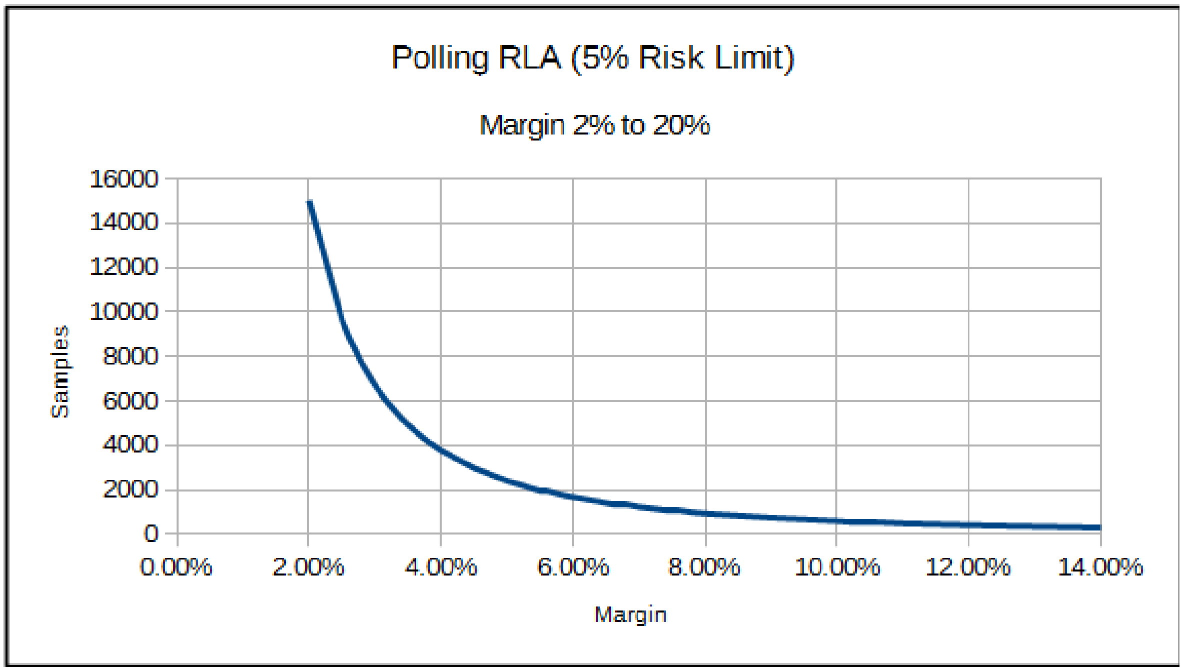


Figure 10: Ballot-sampled Polling RLA with 5% Risk Limit (linear scale)

Although hand-counted paper ballots is viewed as the "gold standard," in practice, humans counting votes on paper ballots produces an error rate between 1% to 2% according to a recent scientific study[11]. So it is better to say that "some other auditing method should be used" at that point rather than asserting that a full manual hand count is the only option.

It is more informative to view these data on a chart with log-log scales, since the 1/x relation then will look like a straight line. Figure 11: Ballot Samples Required for Various Audit Types provides the curve of Figure 9: Samples required for a ballot-sampled polling RLA with 5% risk limit and Figure 10: Ballot-sampled Polling RLA with 5% Risk Limit (linear scale), but also the other curves that will be of interest to us, including some 1% manual tally examples.

[11] Goggin, Bryne, Gilbert, "Post-Election Auditing Effects of Procedure and Ballot Type on Manual Counting Accuracy, Efficiency, and Auditor Satisfaction and Confidence" http://www.copswiki.org/Common/M1725 (2012)

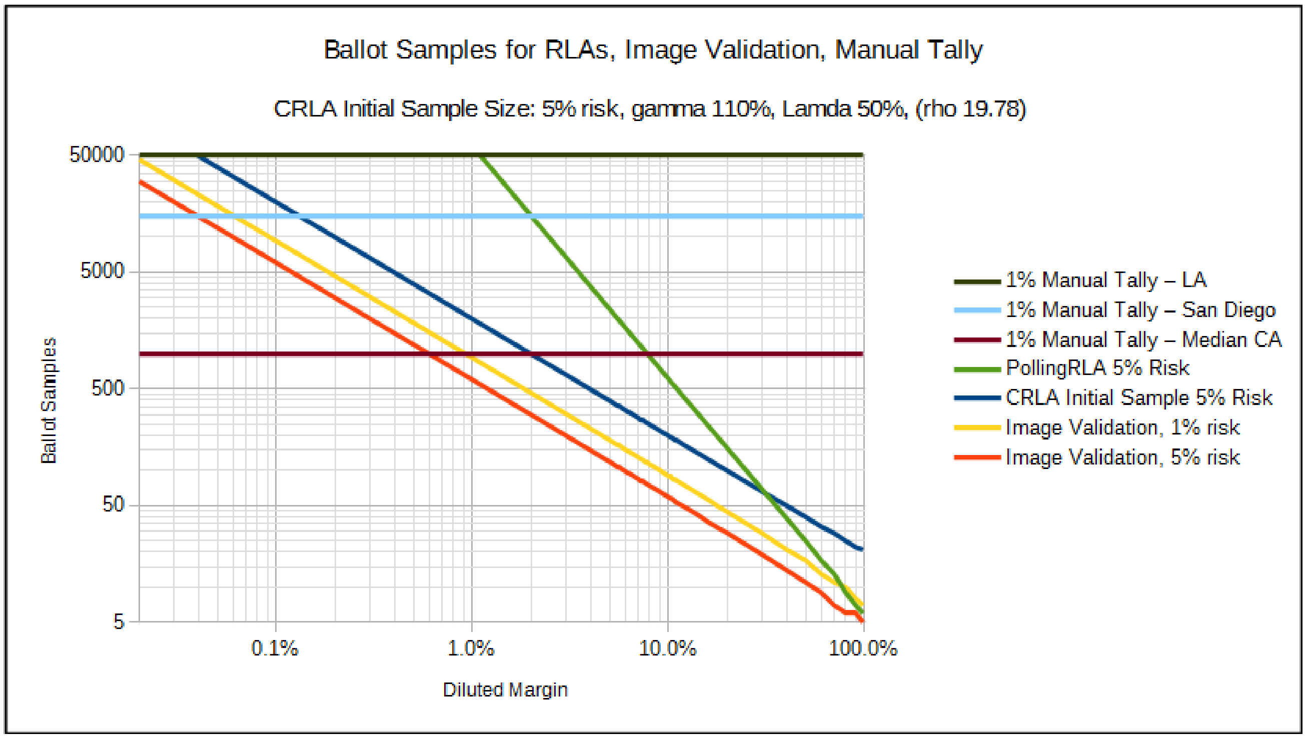


Figure 11: Ballot Samples Required for Various Audit Types

3.2.5 Ballot-sampled "Comparison" RLA

The ballot-sampled "Comparison" RLA (CRLA) is regarded in literature as an improvement to the ballot-sampled "Polling" RLA described above. This version of the audit requires much more information in the form of the CVR for that ballot which can be examined and compared with each physical ballot. The matching ballot must be found among the stored paper ballots, typically through the use of a detailed manifest. As each ballot is examined, if any discrepancies exist between the physical ballot and the CVR, they can be either in the direction to increase the margin claimed by the CVR set or to decrease the margin and perhaps flip the result. The first are called understatements and the second are overstatements, according to the terminology utilized by Lindeman and Stark[12].

For the purposes of analyzing the performance of this audit technique, we will consider the hack already alluded to, that is where a hacker can modify the CVR but cannot affect the ballots, and who wants to modify as few ballots as possible. This can be regarded as the minimum hack that can flip a race.

We can estimate the minimum size of the hack as half the official margin in the race. The hack would have to be at least that large if it changed the winner of the race. Thus if the margin is 3%, then the number of ballots required to be modified to cause a move of 3% is only 1.5%, with two-vote

[12] A Gentle Introduction to Risk-Limiting Audits (AGI) -- http://statistics.berkeley.edu/~stark/Preprints/gentle12.pdf

overstatement to account for moving the votes from one candidate to the other. The hack would likely need to be even larger if the loser was more than 1 vote behind the winner in a true count.

Using the tool provided at https://www.stat.berkeley.edu/~stark/Java/Html/auditTools.htm, the performance of the CRLA is shown for various sizes of hack, with 1.5% rate of two-vote overstatements relating to a 3% minimum resulting margin. Also, shown are the curves for even smaller hacks.

A number of other programs and utilities were utilized, as well as attempting to code the equation directly, and we found that no two of these approaches provided the same numbers. But all had the same characteristics as these curves, where at a specific margin, the count grows without bound, and every margin lower than that would require a full-hand count. (These curves were limited to 35,000, which means that a full hand count is required.)

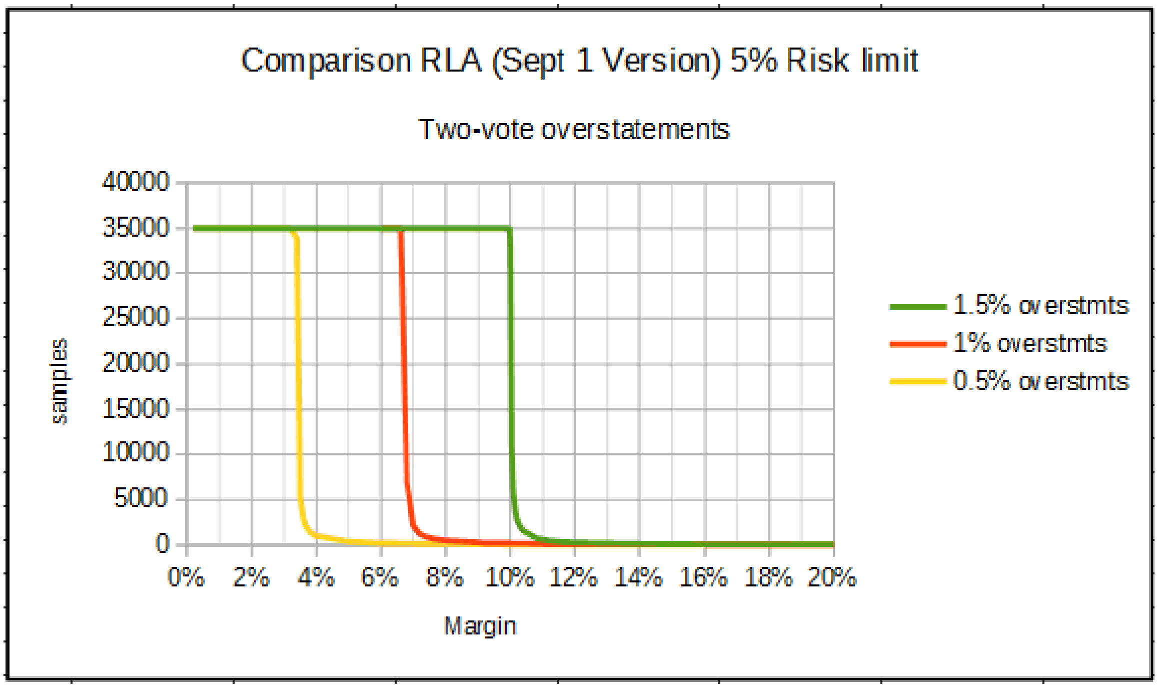


Figure 12: Comparison RLA with 5% risk limit and minimal "flip" hacks

This method can reduce the ballots required to be sampled if it is a landslide victory >>10%, but with a very small 2-vote overstatement hack as described (1.5% of ballots affected), the method results in the need for a different auditing method (the literature calls it the "full manual count") at 10% margin, far before we get to the likely maximum margin caused by that hack (which is 3%).

We should note that if the CRLA encounters no indication of a hack, it will indicate a lower sampling rate and be more economical than the PRLA.

The CRLA method proposed is also not at all simple to explain and utilizes statistics and equations far over the heads of most people. The calculation we need for this analysis, i.e. to discover the sample size based on the margin and the expectation of 1-vote and 2-vote overstatements, is not closed, and requires iteration. A number of programs were evaluated and they did not produce the same results, and even an honest attempt to implementation the equation[13] resulted in different results. This variation and confusion factor is a big problem for those RLA approaches.

Perhaps the easiest way to compare the approaches is to consider the initial step presented in the document "Super-Simple Simultaneous Single-Ballot Risk-Limiting Audits"[14] (on page 4). The first step of the CRLA process is to choose an initial sample of ballots. From that point, the CRLA procedure may require additional samples based on how many 1-vote discrepancies and 2-vote discrepancies exist that are "overstatements," or it may stop after that initial sample.

The CRLA initial sample is "inflated" by two parameters, Gamma and Lambda. The inflation terms only make the process more efficient in terms of short-cutting rapid iterations and additional rounds of random draws. If the CRLA audit finds no 2-vote overstatements and a limited number of 1-vote overstatements, the audit can stop after this first set of random ballots. When these are set to the recommended defaults mentioned in the paper, the curve marked CRLA in " Figure 11: Ballot Samples Required for Various Audit Types" results. If these values are set to not inflate the required samples, the size of the CRLA sample is the same as the sample size required for Image Validation, at the same risk level (Image Validation will be covered shortly).

But if even a single 2-vote overstatement exists, or if there are too many 1-vote overstatements, then the CRLA will escalate and even more ballots will be required. And at a certain point, the CRLA audit procedure calls for "full manual count." In fact, if even one 2-vote overstatement is found in this first set, a full hand count is called for. We will show there are better ways to skin the cat.

3.2.6 Secured and Validated Ballot Images Approach

The other major approach to auditing is based on using secured Ballot Images that are preferably validated. This option does not exist unless ballot images are produced, but this is becoming more common all the time. As we will show, the creation, securing and saving ballot image is a reasonable and prudent goal. Images should be created without lossy compression, in simple formats to avoid have hidden fields that can contain any additional information.

Once you have validated the images as described below, they can be used for just about any audit approach, including a ballot-sampled CRLA, or a 100% independent recount audit. Ballot images should be secured using block-chain style security (i.e. secure hash message digests which are published and signed, this will be explained in the attached technical brief.)

There is a hazard that an insider could modify the CVR and ballot image, but may not have access to the ballots themselves. The only way this can be done is to modify the ballot images before they are secured, and then the CVR is generated from the modified ballot images. We can reduce that possibility

[13] Equation 10 from "Super-Simple Simultaneous Single-Ballot Risk-Limiting Audits" https://www.usenix.org/legacy/events/evtwote10/tech/full_papers/Stark.pdf

[14] ibid page 4

by comparing the ballot images with the paper ballots to confirm they are an accurate representation. This should occur after the ballot images are secured per the Technical Brief – "Block-Chain Style Cybersecurity For Digital Ballot Images" (attached.)

To validate images, the number of ballots that must be sampled is related to the narrowest margin of victory in any race, and the level of certainty we wish to have. The certainty is (1-risk) that we would improperly accept as valid the images when in fact they contain a hack of sufficient size to flip the narrowest race. We assume the most efficient hack, i.e. a 2-vote overstatement (flip of the race by moving the vote from the undesired (winning) candidate to the desired (losing) candidate). This is most efficient hack because the fraudster can modify the outcome by modifying the least number of ballots. Any other modification that alters the result will be less efficient, will require altering more ballots, so those will also be detected if we can detect the most efficient hack.

The question then is: How many ballots must be sampled before at least one of the modified images in the most efficient hack is selected? To calculate this, we must know that the diluted margin is the (race specific margin) * (faction of ballots that include that race).

This is best answered by considering the probability of continuously not selecting a modified image.

Assuming that x% of the ballot images have been modified so they do not match the original ballots, then the probability of not selecting one of those is 1-x. If we do that over and over, then we multiply that probability each time, (1-x)*(1-x)*(1-x)..and so on or (1-x)**n. With every sample, the chance that we will continue to not hit one of the modified images will be reduced accordingly. When (1-x)**n = risk, then n is the minimum number of samples needed.[15]

If one ballot is modified, it could affect the margin by twice that amount, because both candidates could be modified (add one to the lower and subtract one from the winner). Thus, for this application, x = margin/2. First solving for n, and rounding any fraction up.

risk = (1-(margin/2))**n

LOG(risk) = LOG((1-(margin/2))**n)   take log of both sides.

LOG(risk) = n*LOG(1-(margin/2))   pull out exponent

n = CEILING (LOG(risk) / LOG(1-margin/2))   solve for n and round fractions up.   [1]

[15] To explain this for an intuitive understanding, consider a fair coin. How many flips will it take so that 99% of the time, you will see both heads AND tails? The first flip, you will get head or tails with 100% probability, let's assume we get heads. Now, we consider how probable it will be to continue to flip the coin and get heads over and over, until the probability of getting that series is less than 1%. The first flip, we have a 50% chance of getting heads again. Flip again, and the chance is 50% each time, but to get them in series, we multiply the probabilities, 0.50*0.50 =0.25 chance of flipping two more times, then 12.5% (3), 6.25% (4), 3.125% (5), 1.5625% (6), 0.78125% (7). So it took seven more flips. Using Equation [1], n = CEIL (log (0.01) / log (1-.5)) = CEIL (-2/-0.3) = CEIL (6.64) = 7.

This produces the curves shown in " Figure 13: Samples to validate ballot images" using linear scales but the same curves can be seen using log-log scales in " Figure 11: Ballot Samples Required for Various Audit Types."

The best way to perform the validation sampling is to divide the number of ballots to be reviewed by two and select the ballots in two different ways. For the first set, randomly select a CVR, and then using the manifest, look for the matching ballot among the physical ballots and pair it up. For then second set, randomly select physical ballots without using any manifest or computer report, and then attempt to locate the corresponding record in the CVR set using any information available. This approach covers the cases when a ballot is scanned twice (and two records would exist in the CVR set but not in the physical set) or not scanned at all (not in the CVR set but found in the physical ballot set.) Each of those two possibilities can affect the total by only one vote, and so we need to sample them half as often as the flipped-vote case to discover if they are prevalent enough to modify the outcome. [Unfortunately, the CRLA procedures do not perform selection in both ways and so they would miss one class of errors, but they could also implement this as an improvement to those procedures.]

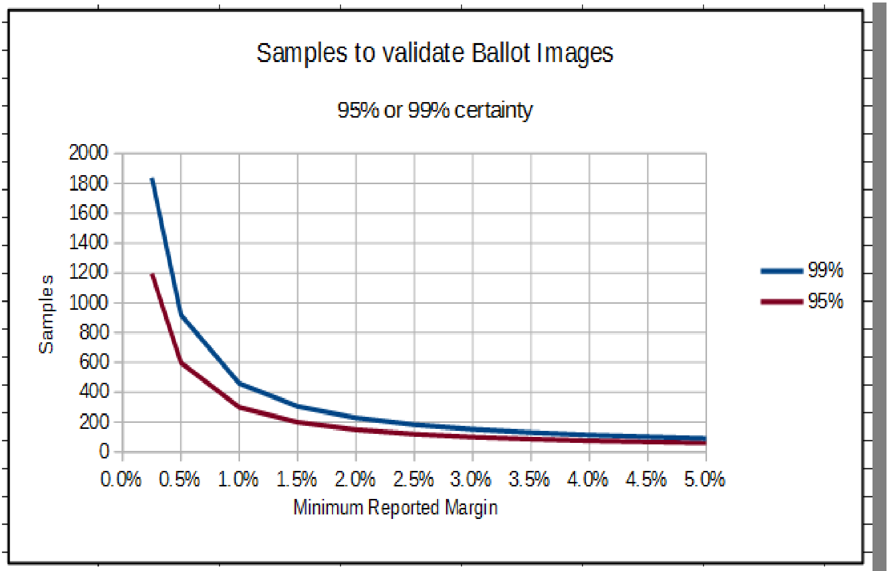


Figure 13: Samples to validate ballot images

The process of selecting and comparing ballot images with physical ballots should be public and documented with recorded side-by-side comparisons, so it will not be feasible for compromised officials to cover up ballots hacked with at least the minimal hack.

It is considered mandatory to have a unique identifier on the physical ballot that can be compared with the image to confirm that the correct image is being matched with the correct ballot. If even a single case is encountered where the ballot image does not match the ballot and the image has been found to be modified, inserted or deleted, then indeed we do have a problem as that should never occur. There is a chance that the image is of the correct ballot but is not a fair representation and then this would indicate likely a scanner or compression issue, or that some ballots are missing from the ballot set (misfed) or a ballot accidentally scanned twice. As the ballot images are compared with the ballots, if the image does not match the ballot, it will need to be determined if the ballot was modified or if the image was modified. If it does appear that the image was modified, this would cause a root-cause analysis to determine how the images were modified and by whom.

For election systems that base their interpretation of the ballots on digital image processing of ballot images, an altered image is an extremely horrendous event to occur, and it means the entire tabulation must be questioned. The policy likely should be that the entire tabulation should be redone from scratch, of course after making corrections so the fraud or severe mistakes could not reccur.

3.2.7 Less work that any other approach

The number of ballots that must be examined to simply validate the ballot set will always be less than the number required by any audit that involves comparison of vote totals, as is the case in the PRLA audit describd above. This is because validating ballot images is much less complex comparison. Ballot images are simply compared and any discrepancy is an indication that the ballot images may be compromised. A single modified image is enough to trigger a ruling that the image set is unreliable. In contrast, the PRLA audits compare a tabulation, which will take many more sample ballots to generate, or compare ballots to get an idea how bad the discrepancy might be.

With the recommended inflation factors, and at the same risk level, the CRLA will require at least twice as many ballots to be physically pulled, and it will only grow from there. Image Validation requires the essentially same number of ballots as the CRLA with no inflation.

3.2.8 Completing the Audit after Images are Validated

Validating ballot images is not a complete auditing procedure. Once the images are validated, an additional auditing approach must be used to then decide if the CVR has been modified.

On the other hand, the downside is that validating images does mean that you have to be able to access both the physical ballot and image for that ballot, similar to the CRLA audit.

The ballots could be reviewed "by hand" by looking over the images, which is comparable with the hand count mentioned, but can be done with many people on the internet in a crowd-sourcing arrangement that can be very satisfying to the public. Or the images could be compared with an auditing program by an independent auditor. More on this later.

4 "Divide and Conquer" and "Test Early, Test Often"

The two key tactics to help test just about anything are encapsulated in the phrases, "Divide and Conquer" and "Test Early, Test Often."

One key shortcoming to the ballot-sampled PRLA or CRLA is that they do not attempt to follow these tactics. The entire ballot set is treated as one random pool, and there are no steps that divide and conquer. And as a result, they do not assist in determining what exactly is wrong -- no diagnostic hints as to what the problem might be is provided.

The approach which first validates ballots images, and then applies another audit technique to the validated samples divides the testing into two phases. When such a split occurs, it provides two benefits. First, each step is simpler and is easier to understand. Secondly, the split allows diagnosis of the problem.

Once the images have been validated, attack vectors based on modifying, adding or subtracting the images have been excluded. After that point, the images can be relied upon (within the risk parameter specified). If the image does not match the CVR, then we know the CVR has been modified, rather than the image modified. Thus, a split in testing provides more information that is useful for diagnosis.

4.1 Non-validated Ballot Images are Still Useful

If secured ballot images are available but they have not been validated by comparing with the paper ballots, can they be used to audit the election? Of course they can, but there is an increased risk. As mentioned, using ballot images splits the problem into two parts. If the ballot images are not validated, that means they have not been compared with the paper, and there is some chance that a hacker may have altered ballot images prior to the creation of the CVR. It will not catch a hack consisting of modifying ballot images after they were created and before they were secured. The CVR will be create based on the altered ballot images, so there will be no difference between the CVR and the Ballot Images in this case. The hack could be detected with any check that compares paper with CVR or Images.

Auditing ballot images means we are comparing them with the CVR set. This will detect any hack that modifies the CVR after the secured ballot images are created, which is largely the entire set of central-tabulator hacks which are possible today. The additional hazards created by creating secured ballot images is far fewer than the hazards which existed without secured ballot images, and this is a net benefit.

Indeed, if the margins are large, it would be quite difficult and hazardous for any compromised insider to modify the ballot images in the tiny window between image creation and image security. With administrative controls, such as improved procedures, this window can be minimized to a point where ballot images can be relied upon for the range of margins where hand counts are not called for, and image validation may be skipped without a huge increase of risk.

4.2 The Open Ballot Initiative

One auditing approach that relies on the existence of secured and validated ballot images is based on the notion that the full set of ballot images can be distributed to several competing groups that will generate their own independent tabulation, creating a full set of cast vote records (CVRs) which can then be easily compared to discover where the disparate groups disagree on the results. This we have called the Open Ballot Initiative (TOBI) because it is based on the concept that once validated, the ballot image evidence can then be easily available for review by anyone.

TOBI suggests that the CVR sets generated by the various third parties are compared ballot by ballot (likely by machine) with the official CVR result to create a result in a standardized format. This is simply comparing a two CVR sets, in essence comparing two large tables, and it is something computers are especially good at (and can also be easily spot-checked manually).

Any ballots where the competing tabulations disagree can be flagged so those can be reviewed in much more detail. Indeed, at some point, the corresponding physical ballots may be consulted and compared with the images as each can provide a secondary check to each other. We don't need to review all the ballots, only the ballots where the different parties disagree. And, each may wish to review paper as well not only to confirm the images that are in question, but also to confirm the set of images as a whole.

This sort of comparison is not a statistical process, it is a rigorous, 100% review of all the ballots by different parties using their own software and algorithms for discerning voter intent. So in the end, the only risk factor is the underlying risk that the images are compromised, and that can be minimized by image validation with a minimum of ballots compared.

What is likely the case, the various competing parties will come up with differing interpretations for some of the ballots. The ballots where voter intent was interpreted differently by the different versions will be some relatively small number of ballots X. If X < half the smallest vote margin, then even if those were all interpreted correctly by exhaustive review, there is no way the that interpretation can change the outcome.

Unlike RLAs, TOBI provides a mechanism for improvements to the voter-intent interpretation heuristics. Over time, the various competing parties -- including the election officials -- can improve their voter-intent algorithms and will have fewer discrepancies. Statistical approaches like PRLA or CRLA do not have this beneficial characteristic.

The fact that TOBI provides the ballot data to other parties provides outsider review of the election so that no insider can cheat. Traditional self-audits which rely on the election officials to also honestly report mistakes require intense scrutiny to insure honest reporting. Independent processing is, in essence, extremely intense scrutiny of the election.

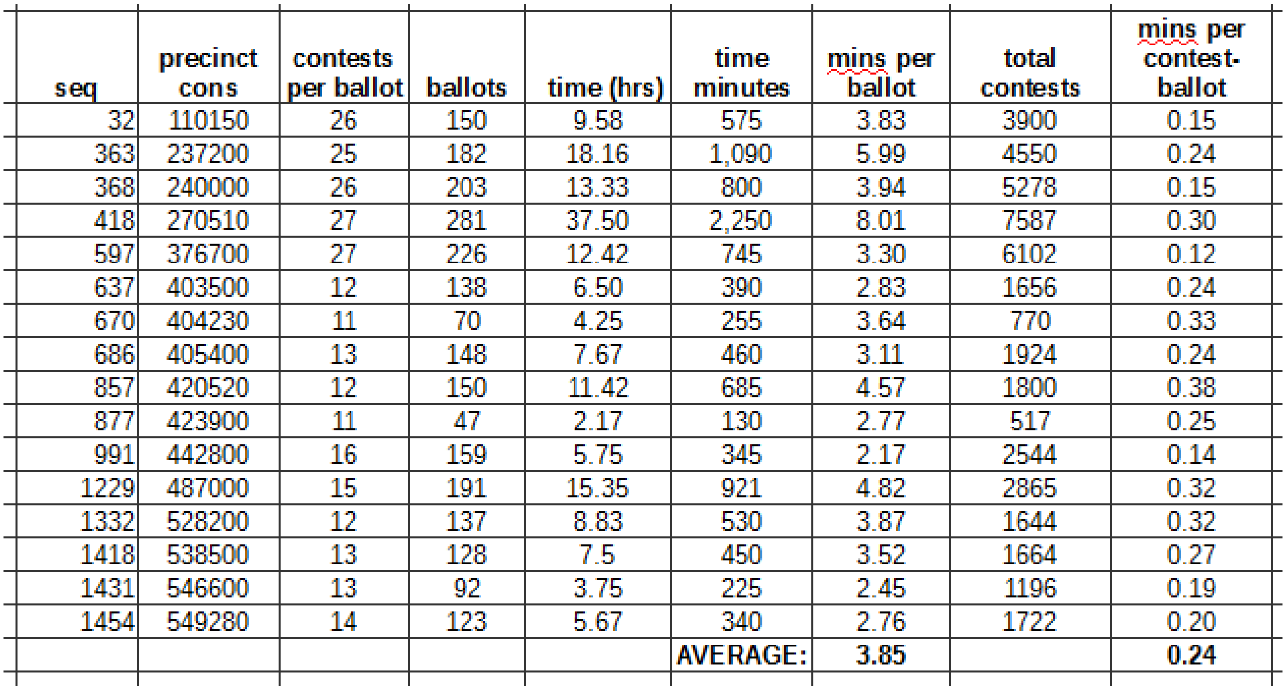
It is interesting to note that if this methodology is used, there is no need to worry about whether the source code of the code providing interpretation of the ballot images is "open-source" or proprietary. The multiple-party comparison process eliminates this from concern, as long as the comparison actually is done, and image validation is done properly and openly.

5 Economic Comparison

One thing that is largely unaddressed is the comparative costs of performing these various types of audits. This section will attempt to understand this.

5.1 Manual Tally Cost

We have good data on the time it takes to perform the 1% Manual Tally as performed in California. Generally, these times do not include the overhead of accessing the batches which are to be random tallied, i.e. precincts in this case, as those are pulled prior to the start of the manual tally process. It is very important for the ballots to be sorted by precinct to implement the 1% manual tally and also to implement any recounts which may be appropriate. If you don't have the ballots sorted by precinct, then you can't easily find the ballots that include the race of interest.



seq	precinct cons	contests per ballot	ballots	time (hrs)	time minutes	mins per ballot	total contests	mins per contest-ballot
32	110150	26	150	9.58	575	3.83	3900	0.15
363	237200	25	182	18.16	1,090	5.99	4550	0.24
368	240000	26	203	13.33	800	3.94	5278	0.15
418	270510	27	281	37.50	2,250	8.01	7587	0.30
597	376700	27	226	12.42	745	3.30	6102	0.12
637	403500	12	138	6.50	390	2.83	1656	0.24
670	404230	11	70	4.25	255	3.64	770	0.33
686	405400	13	148	7.67	460	3.11	1924	0.24
857	420520	12	150	11.42	685	4.57	1800	0.38
877	423900	11	47	2.17	130	2.77	517	0.25
991	442800	16	159	5.75	345	2.17	2544	0.14
1229	487000	15	191	15.35	921	4.82	2865	0.32
1332	528200	12	137	8.83	530	3.87	1644	0.32
1418	538500	13	128	7.5	450	3.52	1664	0.27
1431	546600	13	92	3.75	225	2.45	1196	0.19
1454	549280	14	123	5.67	340	2.76	1722	0.20
					AVERAGE:	**3.85**		**0.24**

Figure 14: Actual data of time cost of 1% manual tally for Polls ballots in San Diego 2016 primary

The manual tally itself can be conducted using the read-and-tally method, with teams of three people, one reader and two talliers. The result of the tally is provided to the supervisor who compares it with the computer result. If the tally result matches the computer, then they stop. Otherwise, it is re-tallied by other teams to determine if the computer result is incorrect. The total time elapsed shown in Figure 14: Actual data of time cost of 1% manual tally for Polls ballots in San Diego 2016 primary of just under 15 seconds per ballot-contest, and nearly 4 minutes per ballot on the average to tally the ballot. These times do not include overhead such as breaks, meal periods, etc. This table only reflects the times for tallying the Polls Ballots, and they do not include any VBM ballots nor provisional ballots.

Studies have shown that for ballots with many races, this read-and-tally method is faster and more accurate than the sort-and-stack method[16]. But if you are doing only one race, the sort and stack method could be faster. To use that, the ballots are sorted into "n" stacks, each representing the vote for each of the "n" candidates. Then, the stacks are counted. The count should total to the total number of ballots.

Practiced teams are important to get these times down. We note that in one precinct it took 37 hours to count it and get it right because it had to be passed to other teams. For those people who are proponents of counting all the ballots in the precinct after the election is over, we remind you that this is a very tedious process and asking the same precinct workers who have worked a 12-hour day should continue to work for at least four and up to 37 hours is asking far too much.

Also, there is another factor. The 1% Manual Tally in CA is a batch-comparison audit. It compares the result with the computer report. If you do not have the computer report to rely on, then it is about twice as costly to do the tally, because you need to tally each precinct at least twice by separate teams to reduce the likelihood of error.

What is surprising about this process is how difficult it is for humans to perform because of the sheer boredom of the work. Yet, these data provide a very good starting point. Unfortunately, we do not have solid data on the other procedures so we will make good-faith estimates.

5.2 PRLA Cost

The PRLA must randomly select physical ballots and tally the result. Thus, the cost of pulling the ballot will be on top of the 4 minute cost to tally any individual ballot. For sake of comparison, we will assume that the ballots can each be accessed within ten minutes[17], and then each ballot tabulated in 4 minutes, for 14 minutes total.

5.3 CRLA Cost

The CRLA requires that the physical ballot be paired up with the CVR record for that ballot and then tallied. So we must find not just a random physical ballot but a specific ballot, and then tally it. Total time probably 20 minutes to get both the physical and CVR ballot and 4 minutes to tally and compare. There is an additional cost to determine if the process must continue or if it can stop. We will include another minute for that, for 25 minutes total.

5.4 Image Verification Cost

The cost to validate images will be similar to the cost included in the CRLA but it will not take 4 minutes to tally the ballot as it need only be compared that they are the exact same ballot. Usually the style of marking in hand-marked ballots can assist in this comparison. We also promote the idea of a "scribble box" where the voter can put in any scribble they wish, preferably not their initials, to make it

[16] Goggin, Bryne, Gilbert, "Post-Election Auditing Effects of Procedure and Ballot Type on Manual Counting Accuracy, Efficiency, and Auditor Satisfaction and Confidence" http://www.copswiki.org/Common/M1725 (2012)

[17] Actual data regarding the time required to access physical ballots is not available as of this publication. Any source of such real data is fondly appreciated.

easier to match up the ballot with the image. Comparing the vote on the ballot need not be part of the comparison process. Thus, we will estimate that IV will take the same time as the CRLA except for the 1 extra minute added for the complexity of the CRLA process.

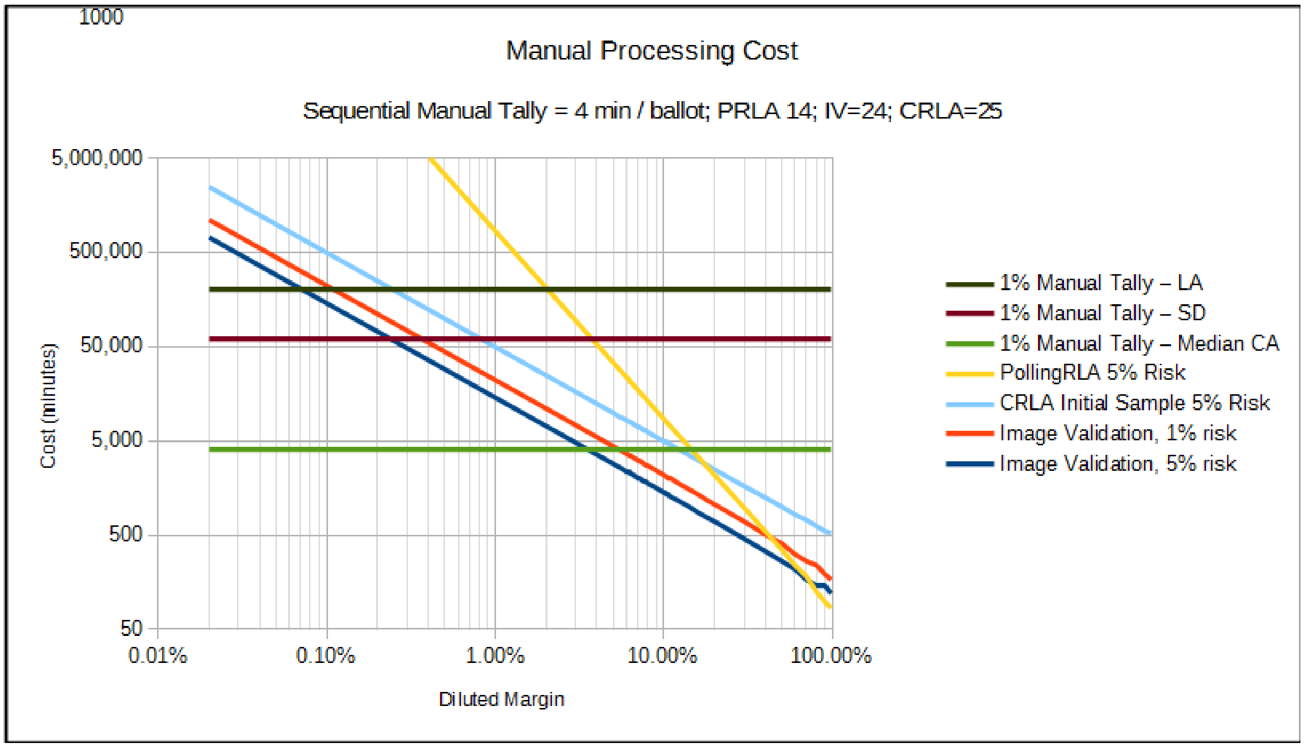


Figure 15: Economic Analysis of Auditing Methods

Figure 15: Economic Analysis of Auditing Methods provides a chart to allow visual comparison of these methods. The main difference in this chart and the earlier similar chart Figure 11: Ballot Samples Required for Various Audit Types is the that the three lines for manual tallying have moved down on the chart with respect to the other curves. The cost to perform the manual tally is based on the actual data for a sequential manual tally. However, if we had just one race that was so close that it needed to be tallied, it might roughly take about the same amount of time as the 1% manual tally for all races. So roughly speaking, we will consider that these levels also represent the cost to tally one race.

We show the median CA county, which has about 100,000 voters, and San Diego, which is the #2 county in the state (and #6 in the nation) with 1.5 million voters, and Los Angeles County, the largest district in the country with about 5 million voters. (The largest districts do not faithfully perform the 1% manual tally on all strata as they tend to omit the Later VBM and Provisional Ballots to reduce the cost.)

6 Combined Strategy

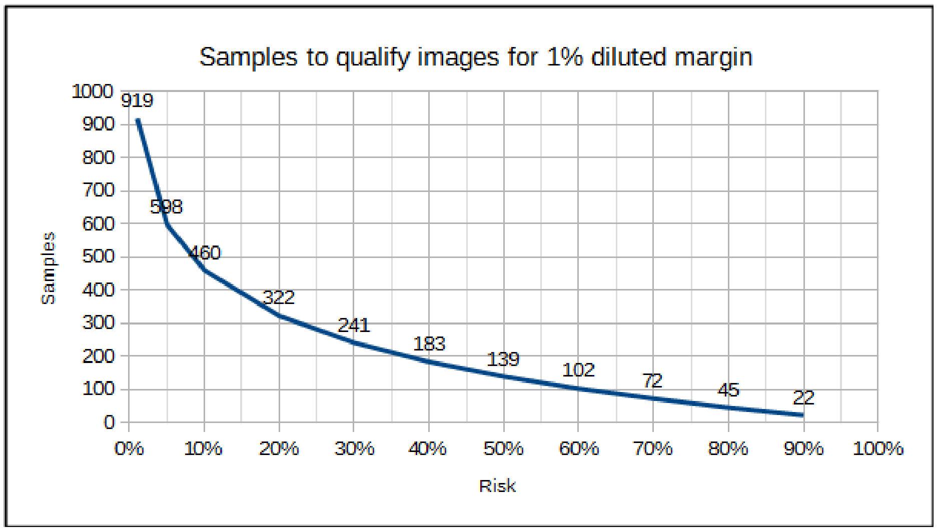
As mentioned earlier, the RLA and Image Verification (IV) curves are the same no matter what size county you have. The number of samples required is only dependent on the margins. Essentially, if the RLA or IV curves are above the line corresponding to the manual count cost for that county (for one race) then, a hand count will be called for. Considering CRLA, half the counties in the state should trigger a full hand count of any race with diluted margin under 10%, because the overall cost to perform the CRLA is higher than just doing the hand count for those races. Using Image Validation, those smaller counties should manually tally races closer than 3% or 4% rather than get involved in any statistical sampling process. If multiple races exist that are close, then CRLA may be economical for slightly tighter margins.

Almost all counties will benefit by the simple rule of manual hand count for any race with 1% margin or narrower. In LA, all hand tally should be used for any race closer than 0.25%.

For the smaller counties, hand tallying races closer than 1% will be more economical than all the work to do anything else. Except for the top 30% of counties (top 18) hand counting races closer than 1% is better than all other methods, as long as only one race is close.

6.1 Leveraging the perceived risk

Statistically speaking, 598 or 919 ballots must be inspected to validate images to deal with a 1% diluted margin, for 5% and 1% risk factors respectively. It can be argued that even though the number should be higher for tighter races, that 919 ballots is enough because the risk is high enough to account for the type of image replacement hack we are looking for, since the actual hack will likely be at least that large to account for the fact that the true loser is probably losing by more than 1 vote, and so the hack will be probably equal to the margin rather than half the margin.



But we have even more leeway with the number of ballots to validate because there is an important psychological effect in play here. If there is even a 50% risk of getting caught, that is too high for any fraudster who will be found out and suffer dire consequences. And the window of vulnerability -- between images being produced and those images being secured -- can be made very narrow through the use of administrative controls. Thus, if we reduce the risk of missing a hack to 50%, we will catch such hacks half the time, that presents a very high risk to any hacker, and thus they will not even try. This can be combined with very careful administrative controls to reduce the critical step between creating the images and securing them, so that the lower confidence of 50% can be tolerated, and that means about 139 ballots should be validated, and that is not a very high cost to pay to provide a real risk to any fraudster who might try.

Further, administrative controls can reduce the risk of a hack due to the difficulty of altering enough ballot images in the tiny window between creating the ballots and securing them.

After the images are secured and then validated, the entire election can be 100% audited. Risk Limiting Audits that are based on a small statistical sample does not have the advantage of the 100% automated audit, and so even though there is some small risk due to the lower than statistically required validation rate, the overall risk is probably lower than the 5% risk which is accepted as high enough for the CRLA process.

6.2 A Note on Random Selection

A key aspect of random selection is that the selection must be a surprise so it is not feasible to anticipate the selection. The usual approach for this is to consider all ballots as one large pool from which random ballots are chosen and then choose using a roll of the dice to set a seed of a public and known random number generator. Although this is mathematically sound, in practice a purely random selection, particularly if the number of ballots in the sample is a small fraction of the total count, will not be uniformly distributed and will not necessarily account for various strata within the ballot universe.

In the election application, there exist identifiable strata in terms of region, processing group, and related to specific machines. If these aspects exist, then the random selection should select from within the existing strata so as to cover all strata uniformly and all machines used uniformly.

Based on our oversight of many random draws at elections across the country, we prefer using dice to select random numbers rather than using the bingo type machines because it is hard to confirm that all the balls are in the machine.

We would rather see a minimum of two random ballots chosen from each precinct in each strata by a roll of the dice rather than relying on the CVR and manifest to choose ballots from the ballot universe at the end. This type of incremental testing should sit well with election officials who may find waiting to the end to start choosing ballots too difficult. Two ballots are chosen so that they can be chosen in the two ways mentioned, starting from the physical ballot and then find the image, or find the image and then locate the physical ballot.

7 Findings and Recommendations

This paper compares various audit approaches and resolves a proposed hybrid approach utilizing secured ballot images which are validated by using very limited paper ballot comparison with ballot images, followed by either a statistical approach or redundant and competitive 100% audits by competing parties.

We offer the following conclusions:

1. **PAPER BALLOTS:** Voter-verified paper ballots are central to any auditing proposal. Ballots should include a unique and random identifier to allow the ballots to be easily matched to ballot images.
2. **ROBUST CHAIN OF CUSTODY:** Chain of custody should include: sealed containers with numbered seals, two persons present at all times with the ballots, and documentation which reflects what exists in the sealed containers, including ballot counts, seal numbers, persons responsible, etc. Ballots should be sorted to the smallest combination of precincts which will require possible recounts of any race.
3. **BALLOT IMAGES:** Equipment which creates high-resolution digital scans of the ballots should be used to create digital image files. Older equipment should be upgraded to use this approach for processing ballots, and all images should be saved.
4. **SECURE THE BALLOT IMAGES:** The digital image files should be secured with block-chain style cybersecurity, as described in the Technical Brief: "Block-Chain Style Cybersecurity For Digital Ballot Images" (attached) or equivalent.
5. **MINIMIZE CRITICAL STEP:** The time between scanning and securing the images should be minimized so as to reduce the risk that images can be modified before they are secured, using at least two workers and write-only media, if available. This is the critical step and if it is tightly regulated, then the lower-confidence level of checking physical ballots with images can be prudently utilized.
6. **BALLOT LEVEL TRACKING**: Every physical ballot should be retrievable and comparable with ballot images. Ballot images and security data should be provided to the public. Voted ballots should be accessible by the public as public records.
7. **VALIDATE IMAGES:** At least two ballots from each batch should be subjected to image validation checking, where first one ballot is chosen at random from the stack of physical ballots and matched to a ballot image, and secondly, a ballot image is chosen to be matched to the physical ballot. These images are checked for fidelity and whether they match each other exactly. The matching should be documented with side-by-side photos. Random numbers are generated by throwing ten-sided dice. An absolute minimum of 139 ballots should be checked

using random selection to validate the images to at least a 50% confidence level for a 1% margin.

If the images are found to have even one example of a ballot which was modified, then the images are invalidated and the entire jurisdiction must rescan their ballots or only use hand-counted paper for any review of tight races. A root-cause analysis should occur to hopefully identify the fraudster responsible.

8. **IMAGES AVAILABLE TO THE PUBLIC**: Ballot images should be made available to the public to copy the files and review the ballots and create their own CVR set, so these can be compared, and those ballot image files can be compared with the digests and signatures to insure no tampering has occurred. (The Open Ballot Initiative, TOBI)
9. **"HAND" RECOUNT ANY CLOSE RACE:** Election jurisdictions with fewer than 100,000 voters should hand count any race which has a margin of less than 1% margin between the winner and runner up, or in multiple candidate races, less than 1% margin between the first excluded candidate and the first included candidate. For jurisdictions larger than 100,000, any race with margins less than 0.5% must be similarly reviewed. Such recounts can use ballot images if they are appropriately secured and validated. Recounts can display the ballot image on the screen so counters with clicking counters can count them, or the race can be recounted using paper ballots using the sort-and-stack method, for best efficiency. Only the tight races must be recounted. This is more cost effective than using any statistical RLA method.
10. **STATISTICAL RLA** (like the Comparison Risk Limit Audit) approaches can be used and will be economical for any jurisdictions larger than 100,000 voters or if margins are greater than 1%. As the statistical RLA is implemented, additional validation of images can occur concurrently and without any additional cost. When the statistics calls for a "full hand count," the district can use images to conduct the review as long as sufficient administrative controls exist to minimize the critical step (Item 5 above) and at least 139 random images have been reviewed, compared, and validated. (The statistical confidence is 50% but will provide significant risk to any hacker, and those hacks will not be attempted.)
11. **FULL 100% THIRD-PARTY AUDITS:** Jurisdictions can elect to provide the secured and validated ballot images to a third party for an official 100% audit of the election. Ballot Image validation is still required but CRLA need not be used.

The most significant conclusion here is that secured and validated ballot images provide a superior approach to auditing. Securing the images is always important because then the physical ballots cannot be modified without detection. Validating the images can be done by sampling at least 139 ballots and comparing the paper with the image, for minimal confidence of 50%, and up to 919 ballots for extremely high confidence of 99%. But if lower confidence levels are acceptable, with the argument that if hackers will not attempt the hack if there is 50% risk of being caught. Then, the rest of the audit is performed by doing a 100% automated recount by a third party using other equipment and personnel, but relying on the same images.

It is hoped that this white paper will help those in the election integrity field to understand the trade-offs among these options.

Ray Lutz, MSEE, has worked in the document imaging industry, contributed to national and international facsimile standards, and is the founder of Citizens' Oversight Projects, a public interest group that has conducted oversight and reviewed election procedures.

More information: http://citizensoversight.org

Contact information:
Ray Lutz; raylutz@citizensoversight.org
619-820-5321

Technical Brief –
Block-Chain Style Cybersecurity For Digital Ballot Images

V1 2018-05-26
V2 2018-07-27
Ray Lutz, CitizensOversight.org
raylutz@citizensoversight.org   619-820-5321

1 Introduction

Digital images of paper ballots is used by “next generation” ballot scanner equipment because they can employ much more advanced image processing to determine the voter-intent. Paper ballots can be destroyed by flood or fire, are expensive to store, and can be easily modified by anyone with a pen. Ballot images, can be inexpensively stored, and once appropriately secured as described below, are impossible to modify without detection.

The method for securing the ballot images should be simple and easily reproduced. It should be utilized as soon as practicable after production of the ballot images. We suggest one simple method below, which is similar to the methods used in crypto-currencies, such as BitCoin, but without the additional complexities not needed in this application.

2 Image Files In a Lot

Digital Ballot images are created by passing the ballot across a linear image sensor or capturing the entire ballot at one time by using a 2-dimensional sensor array, which is the way a typical digital camera does it. Any lossy compression to the images should be disabled, and image sensors should not be blind to any colors.

We assume here that ballots are scanned in "work-units" or "Lots." A Lot can be any convenient group of ballots, perhaps a precinct or batch of vote-by-mail (VBM) ballots.

After the Lot is completed, you will have a set of image files. These files may be simple bit-map format, like .pbm, or some other image file format, such as PDF, TIFF, PNG, etc. There is some valid arguments that the image file should be as simple as possible so there are no hidden crevices where information can be stored as files like PDF, TIFF, PNG, JPG, etc have hidden meta-data which is not immediately apparent. A file format like .PBM is very simple and has no places to hide any information, and once zipped, are still an economical way to store the data.

For example, we will use the ballots published by Dane County, WI, in 2016. Considering “Dunkirk Town Wards 1-6” as the LOT, it has 2,458 images, one for each side of the ballot. A naming convention is used to pair the front and rear images using F and R, as the last letter in the main file name, and the naming convention should also provide the precinct and ballot style. The naming convention used is beyond the scope of this technical brief. Each lot is compressed as a single ZIP archive.

The folder “Dunkirk Town Wards 1-6” looks like this (2458 lines):

```
N0000180000DS01133903640057bb346d664cbF.pbm
N0000180000DS01133903640057bb346d664cbR.pbm
N0000180000DS0113390364006cbd561ff562eF.pbm
N0000180000DS0113390364006cbd561ff562eR.pbm
N0000180000DS0113390364008adffa209c025F.pbm
. . .
W0000190000DS011339036443abde9210ca4f8F.pbm
W0000190000DS01133903644c2980e95731812F.pbm
W0000190000DS01133903646c6f11a32e2f294F.pbm
W0000190000DS01133903646d01e3d0350a2a2F.pbm
W0000190000DS01133903647bc9989f1491128F.pbm
```

Image data for any lot should be available as a ZIP file on the web site of the election district no later than the day they are scanned (if they are scanned after the election day) or if they were scanned prior to or on election day, then they should be published after election night tabulation is completed.

3 Lot Message Digest File

The first step to securing the images is to create a secure message digest for each ballot image file. To generate the secure hash message digest file for all files in this folder, the following command can be used. Here, we will use the MD5 secure hash algorithm[18] which is easily available as the program md5sum.exe[19] for windows, and most Linux distributions include it as a standard utility.

This command

```
md5sum *.* >../LMD_Dunkirk_Town_Wards_1-6.txt
```

Creates the file '`LMD_Dunkirk_Town_Wards_1-6.txt`' in the parent directory, which contains 2458 lines:

```
907b1311c99d6ae2d5a7d688d1aad39c *N0000180000DS01133903640057bb346d664cbF.pbm
8a58dfec42e55c81add0135e90d4217b *N0000180000DS01133903640057bb346d664cbR.pbm
5b65037899b39a9dfa56736f49e21aca *N0000180000DS0113390364006cbd561ff562eF.pbm
09463082ab83a3a8219d482d34ab3e68 *N0000180000DS0113390364006cbd561ff562eR.pbm
2a8d5ce74606b276d3f954d1be756a1b *N0000180000DS0113390364008adffa209c025F.pbm
. . . (snip)

a6de330151abe0e66d483055df595848 *W0000190000DS011339036443abde9210ca4f8F.pbm
f5680d05dc9a8907dca36670e1719682 *W0000190000DS01133903644c2980e95731812F.pbm
302c9221be80913c6daf74ca8eac8641 *W0000190000DS01133903646c6f11a32e2f294F.pbm
f280df7b74759957066f646b9149049e *W0000190000DS01133903646d01e3d0350a2a2F.pbm
cae435c887c74084dd402c9dfb4f75cb *W0000190000DS01133903647bc9989f1491128F.pbm
```

We will note here that the MD5 Secure Hash Digest algorithm, defined in 1991, is not recommended for modern cryptography because there is some remote chance that the digest will be the same for two files that are in fact different, and that it may be able to determine the message from the digest. Because of the nature of this application (the low consequence level if one digest is compromised) it is our

[18] https://en.wikipedia.org/wiki/MD5

[19] http://www.etree.org/md5com.html

opinion that the MD5 algorithm is sufficient and can reduce time costs generating them. But if another (stronger) algorithm is used, it should be expressed as a standard and documented on the website of the election office.

We must realize that here, even the same ballot scanned twice, although the images may appear identical to a human view, the underlying digital data will very very likely be different (and thus result in different message digest values), while ballots that are machine generated may generate the same image file and thus the same digest, no matter how strong the algorithm might be, and yet be considered unique. Hand-marked paper ballots will tend to provide enough uniqueness so no two ballots will be digitally identical. Uniqueness can be added, such as an imprinted ballot ID number.

During the canvass period, there should be a separate ZIP file of the folder of the LMD files, for each day that information is released. The file name should have the date that it is completed.

If the message digests are published and others make copies of what the election office has done, then it is impossible to add or alter any of the image files in any of the lots, nor to add and subtract lots, without detection.

(The method shown here assumes that the scanner equipment does not automatically create the message digest for each image. Modern equipment may create the message digest as each scan is produced, thereby further simplifying the steps involved for the workers and reducing the opportunity for any hacking to occur.)

4 Election Message Digest File

During the election, lots will be incrementally processed, and one LMD file will be added to a folder which contains all the LMDs for the election so far. After each day, an "Election Message Digest" (EMD) file should be created in a similar manner to the command used above, which has one line for each LMD file. That file will provide the Secure Hash Message Digest for each one of the LMD files (which contains, in turn a list of secure message digests and the filename of each image file). In this example, we assumed there are two lots included (so far) in the election, and the EMD file is shown below.

Use this command, where date is filled in with the date of completion.

```
md5sum LMD*.txt >EMD_date.txt
```

Results in this file:

```
ba30f2a4ef65dae434b70d024aa76696 *LMD_Dunkirk_Town_Wards_1-6.txt
1b431d03d49e30129214e54a3a9177dc *LMD-Dunn_Town_Wards_1-7.txt
```

After approximately each day (including any days of scanning prior to the election for early voting) the election district should publish a new EMD_date.txt file (without deleting the prior file). Compared with the prior day, each line in this file will NOT change as the election is completed, lines are added as each lot is processed. Each EMD daily file should be published on the website of the election district (and not published by updating a single file).

It must be emphasized that the EMD files must NOT be coupled or embedded with cast vote record (CVR) data and should be published separately, and prior to any final disclosure of CVR records.

In addition to this file, the election office should produce a cryptographic signature which can be checked using the public key of that election official that the file was indeed produced with their private key. This step will eliminate any ability of the election official of claiming that they did not produce the security information for the ballot images.

5 Standard Publication Format

To encourage and facilitate mechanized interpretation of the files being distributed, they should be published using a file which contains the links to the files available, such as using the RSS 2.0 standard, but more likely using the more recent JSON version of this standard. The exact format of the file and its location to facilitate publication and retrieval is currently under study.

6 Ballot Image QA Procedures

After the ballot images are created and (preferably AFTER being secured using the method described above), the images must be subjected to Quality Assurance (QA) inspection procedures. These procedures are currently under study, but some options are as follows:

1. Check the total number of ballot images matches the number of ballots. A counting scale can be a useful tool to count the ballots based on total weight of the lot.
2. Cut the Lot to find a random ballot. View the ballot image and compare with the physical ballot. (May need to weigh the stacks to determine which ballot is selected if unique identifier is not provided on the ballot.)
3. Select a random ballot in the image set and locate the matching physical ballot.
4. Document the comparison by taking snap shots of the two side-by-side.
5. Evaluate the image quality of the image based on criteria in AIIM TR-34.
6. If a RLA process is used, when any ballot is drawn, it should be compared with the ballot image.
7. If there are ballot images that do not match the source paper ballot, then the ballot images are "invalidated" as a whole. Such a mismatch can occur if the scanner is not equally sensitive to all colors to the same extent, or if lossy compression is incorrectly enabled.

7 Oversight Procedures

Any group providing oversight -- including the Secretary of State -- should download the files from each election district each day. They should check that the EMD file provides the same message digest for each LMD entry compared with files for earlier days. After the image files are available, then oversight groups can check that the image data produces the message digest in each LMD file provided. This will eliminate any risk that files can be added, modified or lots changed.

The Secretary of State should gather up all the MD data from each jurisdiction.

With availability of ballot images, any oversight group can determine the results of the election.

8 Comparison With Official Results

Election officials should prepare a ballot-by-ballot Cast Vote Record CVR with link to the ballot image file, preferably including the same message digest which was provided in the LMD file. Any oversight group that wishes to challenge the results can compare their CVR data with that published by the election district, and provide any specific challenges to official canvass on a ballot-by-ballot basis.

The election office should also provide ballot-styles data, including how the ballot style can be determined by either the image file name or other embedded information, and how the paper ballot can be accessed from secured storage.

More information: http://citizensoversight.org

Contact information:
Ray Lutz; raylutz@citizensoversight.org
619-820-5321

###