Efficiently Auditing Multi-Level Elections

Joshua A. Kroll
Princeton University
Princeton, NJ 08544
kroll@cs.princeton.edu

J. Alex Halderman
University of Michigan
Ann Arbor, MI 48109
jhalderm@eecs.umich.edu

Edward W. Felten
Princeton University
Princeton, NJ 08544
felten@cs.princeton.edu

ABSTRACT

In a multi-level election, voters are divided into groups, an election is held within each group, and some deterministic procedure is used to combine the group results to determine the overall election result. Examples of multi-level elections include U.S. presidential elections and some parliamentary elections (such as those with regional groupings of voters). The results of such an election can hinge on a few votes in one group, while being insensitive to large shifts within other groups. These disparities create opportunities to focus election integrity efforts in the places where they have the highest leverage. We consider how to improve the efficiency of post-election audits, such as those that compare paper ballots to corresponding electronic records, in multi-level elections. We evaluate our proposed solutions using data from past elections.

I. Introduction

A *multi-level election* divides voters into disjoint groups, holds an election within each group, and then applies some deterministic procedure to combine the group results into an overall election result. In this paper, we discuss how to audit multi-level elections efficiently.

An important attribute of multi-level elections is that some ballots may have much more influence than others [1], [2], [3], [4]. For example, in the 2000 U.S. presidential election, a shift of 269 votes in the state of Florida would have changed the national election result, while a shift of 350,000 votes in Texas, or a shift of every vote in the most populous state, California, would not have changed the result. These non-uniformities create opportunities to focus election integrity efforts where they will do the most good. After an election, we can focus our post-election auditing resources to get the highest confidence in the overall election result, at the lowest total cost.

Post-election auditing can help to provide confidence in the integrity of an election by providing evidence that the votes were counted-as-cast. Several electronic election technologies generate redundant copies of ballot data, such as (now widely deployed) optical scan voting systems, in which voters mark paper ballots and scanned images of those ballots are tabulated electronically [5], or systems with a voter-verified paper audit trail, in which voters make a selection electronically and a copy of their selection is printed for review before being dropped automatically into a ballot box [6]. In any system with redundantly stored ballot data (e.g. electronically and on paper), we can audit by comparing the electronic record to the auxiliary record on a per-ballot basis. Generally, the electronic version of the ballot data will be much faster and cheaper to gather and tabulate and the auxiliary record will be much more costly to examine. Thus, we want to minimize the number of auxiliary records that must be examined, while also establishing high confidence that a full examination of all auxiliary records would yield the same election result as the reported electronic result. Efficient post-election auditing relies on examining a subset of the auxiliary records, comparing them to the corresponding electronic records, and relying on statistical arguments to confirm the election result to high statistical confidence.

Much prior work describes efficient approaches to ballot-based auditing in elections with simple majority or plurality rules for determining the election winner from votes cast [7], [8], [9], [10], [11], [12], [13], [14]; this work is the first to consider the case of multi-level elections and how the structure of the election's victory conditions can be used to reduce the total amount of auditing necessary to achieve a certain level of confidence.

Jones gives an overview of the need for and approaches to election auditing [15] and Dopp gives a more complete history of election auditing techniques [16].

Multi-level elections are common. One example is a U.S. presidential election, in which the voters are divided into 51 groups, one for each state.[1] Each state is assigned a certain number of electoral votes. Almost all of the states assign the state's electoral votes to the plurality winner of the state's election. (Two states, Maine and Nebraska, use a different procedure that can divide the state's electoral votes among candidates.) The states' results are combined by summing the electoral votes of each candidate. If one candidate receives a majority of electoral votes, that candidate is the winner. If no candidate receives a majority of electoral votes, then the election result is "undetermined" and the Congress holds a special vote to choose the President.

Another example is a national vote in certain parliamentary systems, where each district chooses a party representative, and representatives from the same party are assumed to act as a single coordinated bloc.[2] In such an election, the result is the identity of the party that holds a majority of seats; or lacking a single party with a majority, the result is the set of minimal coalitions, that is, a set of all of the minimal sets of parties that can form a coalition government. For example, if there are four parties, A, B, C, and D, which have 42, 29, 20, and 9 seats respectively for a total of 100 seats, then the minimal majority coalitions could be formed by parties A and B (71 seats); or by parties A and C (62 seats); or by parties A and D (51 seats); or by parties B, C, and D (58 seats).

Although the practical examples we discuss all determine the overall result by some kind of weighted counting of the individual group results, our theory is much broader than this and can handle any method for combining group results, including, for example, non-monotone systems in which winning more groups can make one's overall result worse. Our theory also extends naturally to handle elections with more than two levels.

The remainder of the paper is structured as follows. In Section II we discuss how to audit multi-level elections. In Section II-A, we work an example showing that considering an election's multi-level structure can reduce auditing costs. In Section II-B and for the rest of the paper, we develop the necessary theory to understand this phenomenon and use it to minimize overall auditing costs. We give an optimal auditing algorithm in Section II-C based on linear programming and in Section II-D, we give an approximation that is sometimes more efficient to compute. In Section II-E and Section II-F, we evaluate these methods using data from several recent elections. We finish by remarking on future work in Section III.

II. AUDITING MULTI-LEVEL ELECTIONS

Post-election auditing is a statistical process for verifying, to some specified level of confidence, that the reported election result is consistent with the available evidence [15]. We assume that there is auxiliary evidence associated with each ballot which can be compared to the reported votes from that ballot, and that the auxiliary evidence is usually unexamined due to cost or time factors [8]. For example, in an optical-scan voting system, the reported results are determined by machine scanners in the polling place, and the auxiliary records are the paper ballots filled out by voters, which can be examined by hand and compared to the machine-reported results. A post-election audit will choose a sample of ballots and compare the chosen ballots with their auxiliary information. If the ballots in the sample are consistent with their auxiliary information, to within a specified tolerance, the audit succeeds; otherwise it fails and further investigation of the election is required.

The purpose of an audit is to reject by statistical means the hypothesis that a full examination of the auxiliary evidence would suggest a different overall election result than the one that was reported. This must be done to some specified level of statistical confidence (sometimes called

[1] For this purpose, the District of Columbia is treated as a state.

[2] This is not a requirement—party members may later defect on particular issues and vote with their opposition. However, we observe that when forming a government, it is especially common for parties to act as blocs (and this is generally expected), making such an assumption reasonable.

the "risk limit" [14]),[3] such as 99%. There is a rich literature on election auditing in one-level popular-vote elections (see [16], [17], [15], [7], [18], [19], [10], [9], [20], [21], [11]). Our method for multi-level auditing could be used with any method that satisfies some general assumptions, as we describe in Section II-C.

Our approach to multi-level auditing will be to assign an auditing responsibility to each group, and then argue that if all groups meet their responsibilities, the overall election result is confirmed in the necessary statistical sense. Because different groups may have a very different impact on the outcome in a multi-level election, we find that auditing to different levels of confidence in different groups can reduce significantly the cost of auditing the entire election to a specified overall level of statistical confidence, $1-\varepsilon$, as we can take advantage of choices about where to direct auditing resources.

Specifically, if the required confidence in the overall result is $1-\varepsilon$, then we will assign group i the responsibility to audit its result to a possibly different confidence level $1-\varepsilon_i$. We will assign the ε_i values such that audit success in every group implies that the overall election result is confirmed with the necessary confidence level.

If an audit in some group fails to confirm the election result, the audit will specify some escalation procedure that aims to determine the correct result in that group. If, ultimately, the election result is changed in some group, it will be necessary to re-evaluate the auditing responsibility assigned to all other groups to ensure that the required confidence level is met. This may necessitate re-auditing or the auditing of additional ballots in some locations if, for example, the auditing responsibility increases in group g' because auditing has changed the reported result in group g. The exact details of escalation will naturally depend on the nature and design of the overall election and the selection procedure that determines the overall result from the outcome in each group.

[3]We stress that, while prior work on election auditing has used the term "risk limit" to describe the acceptable bounds on confidence in the election result, we choose to call this parameter *statistical confidence*, as is done in many other fields. Nonetheless, the concepts are identical: both measure the bounds on the uncertainty in the correctness of the measured election result.

A. *Election Auditing: An Illustrative Example*

To illustrate the mechanics of multilevel election auditing, we will consider the case of presidential elections in the imaginary Republic of Freedonia. Freedonian voters are divided into five districts, District 1 through District 5. They vote directly for candidates for their country's highest office, President. In order to be elected President, a candidate must win a majority of the votes in at least three of the five districts. Thus, Freedonia has a multi-level election: first, candidates must win in each district and second, candidates must win across a majority of districts.

Citizens in Freedonia vote by marking a paper ballot which is scanned by an optical scanning machine that enables fully automated electronic tabulation of the paper ballots. Freedonian election officials wish to verify that the result reported by tabulation of the electronic records is consistent with the paper ballots. They will do this by a statistical procedure designed to verify consistency to 99% statistical confidence, that is, so that any discrepancy between the results will be detected with at least 99% probability. Their goal is to achieve this level of confidence at the lowest cost.

Consider now a specific election in Freedonia between two candidates for President, Alice and Bob. Table I summarizes the results of the election. How should this election be audited?

The most obvious way to audit this election is to conduct a separate audit in each district, to a confidence level of 99% within each district. Because the election within each district uses a simple majority criterion, we can use a standard auditing algorithm from the literature. Calandrino's method [8] would audit 233 ballots in District 1, and 25 ballots in each of Districts 2, 3, and 4, for a total of 308 ballots. (No audit is necessary in District 5 because District 5 did not contribute to Alice's reported victory.) The election result is be confirmed if, for every one of the audited ballots, manual reading of the ballot matches the electronic result reported for the same ballot.

In this case, it is not necessary to audit each individual district to 99% confidence. The reason for this is that Alice was reported as winning four districts when only three were required for victory, so that an incorrect result in only one district could not affect the outcome of the election. In

Candidate	District 1	District 2	District 3	District 4	District 5
Alice	51%	60%	60%	60%	35%
Bob	49%	40%	40%	40%	65%

TABLE I. RESULTS OF THE FREEDONIAN ELECTION, BY DISTRICT.

this case it is sufficient to audit to 90% confidence in Districts 1, 2, 3, and 4. To see why, suppose the election result is incorrect in two districts. If we audited the election 100 times, the audit would detect a discrepancy in the first district in 90 cases, and of the remaining ten cases, a discrepancy would be detected in the second district in nine cases. Only one case out of 100 would go undetected, which yields the required 99% detection rate. Following this procedure, we would audit 116 ballots in District 1 and 13 ballots in each of Districts 2, 3, and 4, for a total of 155 ballots.

Both of the audit strategies we have described so far spend the majority of auditing effort in District 1 (233/308 ballots in the first case, 116/155 ballots in the second case). In general, more ballots must be audited where the election result is close, because only a few miscounted ballots would be sufficient to swing the election and we need to audit more ballots to be confidence that we will randomly choose one of the few miscounted ones. By contrast, when the reported result is not close, auditing fewer ballots yields higher confidence.

This suggests a strategy in which we audit to lower confidence in District 1 and to relatively higher confidence in the other districts. The most extreme version of this strategy does no auditing at all in District 1, and audits to 99% confidence in Districts 2, 3, and 4. The logic of this approach is to establish with 99% confidence that Alice won all of Districts 2, 3, and 4, which is enough to establish that she won the election with 99% confidence, regardless of the accuracy of the reported District 1 results. In this approach we audit 25 ballots in each of Districts 2, 3, and 4, for a total of 75 ballots.

The Freedonia example shows that clever multilevel auditing strategies can reduce substantially the cost of auditing without reducing confidence in the result. It also illustrates some of the strategies that are possible. The results of analyzing this example are summarized in Table II.

The remainder of this paper presents a general mathematical theory for finding the lowest-cost strategy for auditing the result of any election conducted under a multi-level election procedure.

B. Basic Theory of Multi-Level Auditing

Intuitively, if the result of a multi-level election is incorrect, then it must be the case that the within-group result is incorrect for a sufficiently large set of the constituent groups. We define a *flipset* to be a set of groups such that changing the election results in all of these groups would have changed the overall election result. For example, in a U.S. presidential election, a flipset is a set of states which, if they all changed their results, would collectively change the total electoral college winner. We will say that F is a *minimal flipset* if F is a flipset but no proper (i.e., smaller) subset of F is a flipset. If F is a flipset, then there is some minimal flipset F^* such that $F^* \subseteq F$.

It is easy to show that if α_i are chosen so that for every minimal flipset F, $\sum_{i\in F} \alpha_i \geq 1$, and if the reported result in every group i is confirmed to confidence level $1-\varepsilon^{\alpha_i}$, then the overall election result is confirmed to confidence level $1-\varepsilon$. The intuition behind the proof is that if the reported overall election result is wrong, then there must be some minimal flipset F^* such that the reported group results are wrong for every group in F^*. The probability that the audits will fail to notice anything wrong anywhere in F^* is $\prod_{i\in F^*} \varepsilon^{\alpha_i} = \varepsilon^{\sum_{i\in F^*} \alpha_i}$ which by assumption is at most ε.

Cox gives a taxonomy of voting systems [22]. Our methods apply to any voting system which partitions voters into disjoint groups and holds an election in each group, subject to the constraint that the outcome at each level above the first is determined simply from the win or loss condition at the previous level (and not properties specific to the voting system used, such as vote counts).[4]

C. Optimal Auditing for Multi-Level Elections

We now turn to the question of how to minimize the cost of auditing a multi-level election.

[4] *Mixed member proportional* systems, such as the one used for parliamentary elections in Germany, do not have this property.

District	99% Confidence/District	90% Confidence/District	Optimal
District 1	233	116	0
District 2	25	13	25
District 3	25	13	25
District 4	25	13	25
Total	308	155	75

TABLE II. COST OF AUDITING THE FREEDONIAN ELECTION, IN TERMS OF NUMBER OF BALLOTS EXAMINED, BY STRATEGY EMPLOYED.

We allow the use of any known auditing scheme within each group. Our only assumption is that the expected cost C_i of auditing group i to confidence level $1-\varepsilon^{\alpha_i}$ can be expressed as $C_i = t_i \cdot \alpha_i$ for a group-specific coefficient t_i. Because t_i is the *expected* cost coefficient, our model can accommodate underlying audit methods that make adaptive decisions as to when to stop auditing, as well as schemes that have different audit costs for different ballots within a group.

We start by observing that many auditing schemes have a linear cost property, so that the expected cost of auditing a group of ballots to confidence level $1-\varepsilon^{\alpha_i}$ is proportional to α_i, with the constant of proportionality depending on the auditing scheme and the number and distribution of ballots. This constant will typically differ from group to group.

To see why linearity is a natural relation, consider that many auditing algorithms operate by performing a test (such as examining one ballot) and repeating the test, with an independent random selection, as many times as necessary until a desired confidence level is reached. If one test costs C_0 and achieves confidence $1-\varepsilon^{\alpha_0}$, then repeating the test k times (and failing if any of the k instances fails) will yield confidence $1-\varepsilon^{k\alpha_0}$ at expected cost kC_0, which satisfies the linear cost property.[5]

In the remainder of the paper, we will assume an audit scheme that has the linear cost property, that is, that the *expected* cost of auditing, *within each group* is linear *in the parameter* α_i. For schemes whose cost functions are approximately linear, our algorithm will yield a strategy that meets the required confidence level, and with cost that will typically be close to optimal. Finding the optimal-cost solution for nonlinear cost scheme will be more expensive, requiring nonlinear optimization.

If an audit scheme does not have the linear cost property, it would be fairly easy to apply our techniques using nonlinear optimization methods such as hill climbing, especially since the number of variables (i.e. the number of groups in the first-level partition of voters) is usually very small (e.g. in the U.S. Presidential election, there are 51 partitions at the lowest level). One could also approximate the cost function linearly near a proposed solution, which would lead to a correct solution (in the sense that the audit would function to guarantee the specified statistical confidence), although not necessarily a cost-optimal solution.

Because we assume the cost is linear in the α_i, we can use linear programming to find values for the α_i that minimize the total cost, subject to the constraints discussed above. For each minimal flipset F, we will have a linear constraint $\sum_{i\in F} \alpha_i \geq 1$. This will give us the optimal (lowest-cost) auditing procedure that achieves the required confidence level.

In Appendix A, we prove that two well-known ballot-based auditing methods, the Machine-assisted Election Auditing algorithm by Calandrino *et al.* [8] and the Secrecy-preserving Ballot-level Audit (SOBA) of Benaloh *et al.* [13], have the linear cost property required by our scheme.

D. Score-Based Auditing Method

In some cases, it may be difficult or inconvenient to use linear programming to find the optimal assignment of α values. As an alternative, we can approximate the solution using a score-based method that provides the required level of confidence but not a guarantee of minimal cost. To do this, we choose some method of assigning a non-negative numerical score to each group. If group i has score s_i, and if we can show that

[5]It is possible to scale to a non-integer multiple of the original α_0 and C_0 by probabilistic interpolation: if k is an integer and $0 \leq f < 1$, then an algorithm that performs the base audit k times, then with probability f performs the base audit one more time, will be linear, giving confidence $1-\varepsilon^{(k+f)\alpha_0}$ at expected cost $(k+f)C_0$.

any minimal flipset must have total score at least s_*, then we can assign $\alpha_i = \min(1, \frac{s_i}{s_*})$. (Groups that do not appear in any minimal flipset can be assigned $\alpha_i = 0$.) It is easy to show that this will be feasible, in the sense that the α values in any minimal flipset will sum to at least 1.

As an example, in a U.S. electoral vote election, we could assign each state a score equal to its number of electoral votes. If the electoral vote margin is M (that is, if at least M electoral votes would have to flip to change the election result), then it is easy to see that any minimal flipset must have total score at least $s_* = M$. Applying the score-based auditing method, a state i having e_i electoral votes gets $\alpha_i = \min(1, \frac{e_i}{M})$. The intuition is that a state's fair share of the "α burden" is proportional to its number of electoral votes.

As a refinement, we can assign $\alpha = 0$ for a subset of groups, presumably because auditing is especially expensive for these groups. We can choose a set D of groups to "drop", such that $M_D = \sum_{i \in D} \alpha_i$ is less than M. Then for every $i \in D$ we set $\alpha_i = 0$; and for every i not in D we set $\alpha_i = \min(1, \frac{e_i}{M - M_D})$. The intuition is that we don't bother to audit the groups in D, but we increase the auditing burden proportionally in the remaining groups to ensure that the total α in every minimal flipset is still large enough.

These score-based methods are likely to be useful when the number of minimal flipsets is very large. For example, in the 2008 U.S. presidential election, there are 79 841 552 minimal flipsets. Rather than enumerating them and solving a large linear programming problem, the score-based method can yield a much faster solution that we conjecture will often be close to optimal.

We observe that our score-based method is similar to the method introduced by Aslam, Popa, and Rivest [18]. That method divides votes into groups (typically precincts), but assumes that vote totals in each group are always summed to get the overall election result. We allow arbitrary aggregation rules across groups, subject to the constraint that the rules must only consider the win/loss outcome in each group. Additionally, Aslam *et al.* assume that auditing within a group is all-or-nothing: either a group is audited to 100% confidence or not at all. We admit different levels of auditing leading to different confidence intervals. Finally, our main method accounts for the bin-packing issues associated with allowing mixed confidence levels across groups, while the score-based method and the Aslam *et al.* method both ignore these issues for the benefit of ease of computation.[6]

E. Application to U.S. Presidential Elections

To illustrate the use of multi-level auditing, we can apply our method to the U.S. Presidential election from 2000 through 2012, as summarized in Figure III. As an example, the 2012 election was won by Barack Obama with 332 electoral votes, over Mitt Romney's 206. For this election, a minimal flipset would be any minimal set of states that were won by Obama and add up to at least 63 electoral votes. This calculation assumes that the expected per-ballot cost of auditing is equal in all states, so that all $c_i = 1$.

The 2000 election was very close, so there are few flipsets. Any one of the states won by Bush forms a singleton minimal flipset, so the optimal auditing strategy requires that each of these thirty states be audited to confidence level 99%. At the other extreme, the 2008 election had a larger margin of 96 electoral votes, leading to roughly 80 million minimal flipsets. Our linear program solver ran out of memory on this example, so we show a cost only for the score-based method.

F. Application to the 2010 UK Parliamentary Election

As another illustration, we applied our methods to the 2010 parliamentary election in the United Kingdom. Separate plurality elections were held in each of 565 districts. In total, members of twelve parties won seats, with the Conservative party winning 306 seats, the Labour party 258, the Liberal Democrats 57, and smaller parties winning 8, 6, 5, 3, 3, 1, 1, 1, and 1 seats, respectively. For purposes of auditing, we assume that each party's members will vote as a bloc. Since no party has a majority, a coalition of parties holding at least 326 seats in total is required to govern. We considered the set of possible governing coalitions to be the election result.

[6]While Aslam *et al.* consider linear programming as an optimal solution and give a linear program formulation of their method, they dismiss the result as necessarily too costly and complex to calculate.

Year	Electoral Vote Margin	Num. Minimal Flip Sets	Expected Number of Ballots Audited ($\varepsilon = 0.01$)			
			State-by-State	Score-Based	Score w/ Drop	Optimal (LP)
2012	63	872,775	2691.9	475.2	421.1	421.1
2008	96	79,841,552	7705.8	430.6	220.7	·
2004	17	5896	5262.6	1239.9	1239.9	1183.6
2000	2	30	64145.9	51651.1	51651.1	51651.1

TABLE III. AUDITING REQUIREMENTS FOR U.S. PRESIDENTIAL ELECTIONS 2000-2012. EXPECTED AUDITING COSTS (ASSUMING UNIT COST PER BALLOT AUDITED) REQUIRED TO ACHIEVE AN OVERALL CONFIDENCE OF 99% ($\varepsilon = 0.01$) VS. ELECTORAL VOTE MARGIN AND NUMBER OF MINIMAL FLIP SETS, AS CALCULATED USING THE OPTIMAL LINEAR PROGRAMMING METHOD, THE SCORE-BASED METHOD, THE SCORE-BASED METHODS WITH DROPS, AND A METHOD WHICH CONSIDERS AUDITING TO 99% CONFIDENCE IN EACH STATE SEPARATELY FOR THE U.S. PRESIDENTIAL ELECTIONS IN YEARS 2000-2012. FOR THE 2008 ELECTION, OUR LINEAR PROGRAM SOLVER RAN OUT OF MEMORY, SO WE SHOW ONLY THE SCORE-BASED RESULTS.

Given these assumptions, there turn out to be many possible governing coalitions that control the bare minimum number of seats. Every party can participate in such a minimum-size governing coalition. As a result, for every seat there is a minimal flipset containing only that seat, so that every seat i must be assigned $\alpha_i = 1$. Auditing to a 99% confidence level requires examining an expected 98384 ballots.

The amount of auditing required might have been much less had the election come out differently. For example, if the three major parties had gotten 256, 208, and 157 seats, and the minor parties were unchanged, then there would be only three minimal coalitions, consisting of all pairs of major parties. In this scenario the minor parties do not matter, and the smallest minimal governing coalition is a Labour-LibDem coalition with 365 seats. In this scenario, every minimal flipset involving Conservative seats contains at least 88 seats, and every minimal flipset containing Labour or LibDem seats contains at least 40 seats. Therefore we can assign every Conservative seat $\alpha_i = \frac{1}{88}$, every Labour and LibDem seat $\alpha_i = \frac{1}{40}$, and every minor party seat $\alpha_i = 0$. This would correspond to auditing every Conservative seat to a confidence level of only 0.05, and every Labour and LibDem seat to a confidence level of only 0.11. For most seats, the expected number of audited ballots would be less than one.

In general, an approach to auditing parliamentary coalition elections of this type is to compute all of the minimal coalitions (i.e., all coalitions which do not have a proper subset that is a coalition), and then to compute, for each party, the coalition containing that party which contains the smallest number of seats. Let x_i be the size (in seats) of the smallest coalition containing party i, and let x^* be the minimum number of seats needed to form a coalition (i.e., one more than half of the seats). If $w(i)$ denotes the party that won seat i, we can assign the score

$$s_i = \frac{1}{1 + x_{w(i)} - x^*}.$$

It is easy to show that any minimal flipset must have total score at least 1, so we can assign $\alpha_i = s_i$. As an additional optimization, we could consider "dropping" some seats in order to reduce the total auditing cost.

III. CONCLUSION

We introduce a novel auditing technique for examining confidence in and the integrity of real-world multi-level election systems such as the electoral college in the U.S. presidential election or coalition parliament systems in many countries.

Specifically, we describe a method for ballot-based auditing which uses the structure of the multi-level election to reduce the total amount of auditing necessary to achieve full confidence in the overall election result. We show how to use the particular structure of multi-level elections to reduce or ignore the auditing of some subgroups, reducing the cost of auditing while maintaining a defined level of overall confidence. We show both a cost-optimal approach to auditing the overall election to a specific level of statistical confidence $1 - \varepsilon^{\alpha}$ and also a score-based approximation that yields an easily computable correct, but not necessarily cost-optimal, audit strategy. We evaluate this method on real election data from the U.S. and the U.K. and show that it can significantly reduce auditing costs (in our U.S. presidential election examples, costs using our strategy were between 15.2% and 80.5% of a strategy that was independent of the election's multi-level structure; in an example drawn from the U.K. Parliamentary election in 2010 (in which the results were highly split, allowing for many different possible coalitions), auditing to 99% confidence requires a modest cost of examining just under 100,000 ballots).

As future work, we intend to apply our frameworks to more elections and more types of election systems around the world. For example, we have only considered concretely elections where the first level in the multi-level system is decided by a majority or plurality vote. However, our results generalize readily to any selection algorithm, and we intend to consider such alternative systems in detail. For example, certain kinds of *mixed member proportional* systems (and related systems, such as those used in Germany), are not multi-level in the way we have defined. However, we believe our methods can be generalized to include such systems. We are also further refining our algorithms for determining optimal audit costs and seek to find more efficient algorithms, which are still provably cost-optimal.

APPENDIX

We give two concrete examples of multi-level election auditing using ballot-based auditing algorithms that satisfy the linear cost property. We assume in these examples for simplicity that the within-group elections are decided by a simple plurality or majority[7] and that auditing k ballots in group i has expected cost $k\ell_i$ for some group-specific expected per-ballot examination cost ℓ_i.

1) Example: Calandrino's Ballot-Based Audit: First, we give an example using election auditing algorithm of Calandrino et al. [8], which obeys the linear cost model.

Consider a group i with expected per-ballot auditing cost ℓ_i and an assigned responsibility to audit to confidence level $1-\varepsilon^{\alpha_i}$. Let m_i be the victory margin of the winning candidate. In a plurality election, $m_i = \frac{v_i^1 - v_i^2}{2}$ where v_i^1 is the winning candidate's vote count and v_i^2 is the second-place candidate's vote count. In a majority election with a winner, $m_i = v_i^1 - \frac{v_i}{2}$ where v_i is the total number of votes cast in the group. In order for the declared group winner to be wrong, at least m_i of the votes cast for the group's winning candidate must be defective, so that at least a fraction m_i/v_i^1 of the declared winner's votes must be defective. It follows that auditing n_i of the ballots cast for the group winner, without finding an error, will confirm the accuracy of the group winner with confidence level $1-(1-m_i/v_i^1)^{n_i}$, so that we can achieve the desired confidence level $1-\varepsilon^{\alpha_i}$ by setting

$$n_i = \frac{\alpha_i \log \varepsilon}{\log\left(1 - \frac{m_i}{v_i^1}\right)}.$$

(If the resulting n_i is not an integer, we can interpolate: if $n_i = k + f$ for integer k and $0 \le f < 1$, we choose k ballots with probability $1-f$ and $k+1$ ballots with probability f. Then the expected number of ballots chosen is equal to $k + f = n_i$ and the other necessary properties hold.)

Applying the same argument to all groups, we see that the total auditing cost will be

$$C = \sum_i \ell_i n_i = \sum_i \ell_i \frac{\alpha_i \log \varepsilon}{\log\left(1 - \frac{m_i}{v_i^1}\right)}.$$

Setting

$$\ell_i' = \frac{\ell_i \log \varepsilon}{\log\left(1 - \frac{m_i}{v_i^1}\right)}$$

the cost becomes

$$C = \sum_i \ell_i' \alpha_i.$$

This is consistent with the linear-cost model.

2) Example: SOBA: To emphasize that any auditing algorithm with linear expected cost could be substituted, without changing our basic analysis, we provide a second example using SOBA [13], a modern risk-limiting audit method, which also has the necessary property that the expected cost of auditing within each group i is linear in the parameter α_i.

We assume as before that subgroup elections are decided using simple plurality or majority first-past-the-post rules and that the election in each subgroup yields a well-defined result.

Consider now group i with expected per-ballot auditing cost ℓ_i and assigned responsibility to audit to confidence level $1-\varepsilon^{\alpha_i}$. Say that the winning candidate has margin m_i Then the SOBA "diluted margin" will be m_i/N_i where N_i is the number of ballots cast in group i. That means that, given numerical parameters λ and γ,

[7] A majority election might have no winner; in that case we consider the result to be $\perp$. To simplify the exposition, we will assume in the main text that $\perp$ is not the declared result in any group, although our algorithms can easily be extended to cover that case.

the "error tolerance" and "error bound inflator", respectively, the number of ballots audited in the first round of SOBA is:

$$n_i^0 = \frac{\alpha_i}{\frac{1}{2\gamma} + \lambda \log\left(1 - \frac{1}{2\gamma}\right)}$$

SOBA proceeds by adding ballots to this sample until a specific confidence threshold is achieved. The expected additional cost from repeating the audit is negligible, scaling as C^{-2m} where C is a constant derived from the margin of victory and m is the number of misstated votes discovered [12].

The total cost C is obtained by summing over all groups gives:

$$C = \sum_i \ell_i n_i = \sum_i \frac{\alpha_i}{\frac{1}{2\gamma} + \lambda \log\left(1 - \frac{1}{2\gamma}\right)}$$

And setting:

$$\ell_i' = \frac{\ell_i}{\frac{1}{2\gamma} + \lambda \log\left(1 - \frac{1}{2\gamma}\right)}$$

we again obtain (consistent with the linear-cost model):

$$C = \sum_i \ell_i' \alpha_i.$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for thoughtful comments on an earlier version of this paper submitted to JETS Volume 1, Issue 1.

REFERENCES

[1] J. F. Banzhaf III, "Weighted voting doesn't work: A mathematical analysis," *Rutgers L. Rev.*, vol. 19, p. 317, 1964.

[2] L. S. Penrose, "The elementary statistics of majority voting," *Journal of the Royal Statistical Society*, vol. 109, no. 1, pp. 53–57, 1946.

[3] L. S. Shapley and M. Shubik, "A method for evaluating the distribution of power in a committee system," *American Political Science Review*, vol. 48, no. 03, pp. 787–792, 1954.

[4] K. J. Arrow, "A difficulty in the concept of social welfare," *The Journal of Political Economy*, vol. 58, no. 4, pp. 328–346, 1950.

[5] D. W. Jones, "On optical mark-sense scanning," in *Towards Trustworthy Elections*. Springer, 2010, pp. 175–190.

[6] R. T. Mercuri, "Electronic vote tabulation checks and balances," Ph.D. dissertation, University of Pennsylvania, 2001.

[7] J. A. Aslam, R. A. Popa, and R. L. Rivest, "On estimating the size and confidence of a statistical audit," in *USENIX/ACCURATE Electronic Voting Technology Workshop (EVT'07)*, 2007.

[8] J. Calandrino, J. Halderman, and E. Felten, "Machine-assisted election auditing," in *USENIX/ACCURATE Workshop on Electronic Voting Technology (EVT'07)*, 2007.

[9] P. B. Stark, "Conservative statistical post-election audits," *Annals of Applied Statistics*, vol. 2, pp. 550–581, 2008.

[10] J. L. Hall, P. B. Stark, L. W. Miratrix, M. Briones, E. Ginnold, F. Oakley, M. Peaden, G. Pellerin, T. Stanionis, and T. Webber, "Implementing risk-limiting post-election audits in california," 2009.

[11] S. Checkoway, A. Sarwate, and H. Shacham, "Single-ballot risk-limiting audits using convex optimization," in *Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT'10)*, 2010.

[12] P. B. Stark, "Super-simple simultaneous single-ballot risk-limiting audits," in *Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT'10)*, 2010.

[13] J. Benaloh, D. Jones, E. L. Lazarus, M. Lindeman, and P. B. Stark, "Soba: secrecy-preserving observable ballot-level audit," in *Electronic Voting Technology Workship/Workshop on Trustworthy Elections (EVT/WOTE'11)*, 2011.

[14] J. Bretschneider, S. Flaherty, S. Goodman, M. Halvorson, R. Johnston, M. Lindeman, R. L. Rivest, P. Smith, and P. B. Stark, "Risk-limiting post-election audits: Why and how," Oct. 2012.

[15] D. W. Jones, "Auditing elections," *Communications of the ACM*, vol. 47, no. 10, pp. 46–50, Oct. 2004.

[16] K. Dopp, "History of confidence election auditing development (1975 to 2008) & overview of election auditing fundamentals," *National Election Data Archive*, 2008.

[17] W. R. Mebane, Jr., J. S. Sekhon, and J. Wand, "Detecting and correcting election irregularities," Stanford University, Tech. Rep., Oct. 2003. [Online]. Available: http://wand.stanford.edu/research/detecting.pdf

[18] J. A. Aslam, R. A. Popa, and R. L. Rivest, "On auditing elections when precincts have different sizes," in *USENIX/ACCURATE Electronic Voting Technology Workshop (EVT'08)*, 2008.

[19] P. B. Stark, "A sharper discrepancy measure for post-election audits," *The Annals of Applied Statistics*, vol. 2, pp. 982–985, Nov. 2008.

[20] A. D. Sarwate, S. Checkoway, and H. Shacham, "Risk-limiting audits and the margin of victory in nonplurality elections," *Statistics, Politics, and Policy*, vol. 3, no. 3, pp. 29–64, 2013.

[21] R. L. Rivest and E. Shen, "A bayesian method for auditing elections," in *Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE'12)*, 2012.

[22] G. W. Cox, "Centripetal and centrifugal incentives in electoral systems," *American Journal of Political Science*, pp. 903–935, 1990.