Election Security with Risk-Limiting Audits

Oliver Broadrick[1] Poorvi L. Vora[1] Filip Zagórski[2,3]

[1]Department of Computer Science, The George Washington University [2]Wroclaw University of Science and Technology [3]Votifica

THE GEORGE WASHINGTON UNIVERSITY WASHINGTON, DC

THE GEORGE WASHINGTON UNIVERSITY WASHINGTON, DC

Risk-Limiting Audits

- In election security, a risk-limiting audit (RLA) draws trusted paper ballots in rounds, stopping if a rigorous statistical criterion is satisfied, or proceeding to a full hand count. If the announced outcome of the election is erroneous, an RLA will detect the error with high, predetermined minimum probability.
- An audit $\mathcal{A}$ takes a sample of ballots X as input and gives as output either (1) $Correct$: the audit is complete, or (2) $Uncertain$: continue the audit.
- The maximum risk R of audit $\mathcal{A}$ with sample $X \in \{0,1\}^*$ drawn from the true underlying distribution of ballots is

$$R(\mathcal{A}) = \Pr[\mathcal{A}(X) = Correct \mid H_0],$$

where H_0 is the null hypothesis: the true underlying election is a tie.

- An audit $\mathcal{A}$ is a Risk-Limiting Audit with risk limit α iff

$$R(\mathcal{A}) \leq \alpha.$$

BRAVO and Minerva

- Existing RLAs include BRAVO (which can be implemented as End-of-Round (EoR) BRAVO or Selection-Ordered (SO) BRAVO) and MINERVA. In a first round with a large sample size (probabilty of stopping at least 90%), MINERVA is known to require roughly half as many ballots as EoR BRAVO.
- We provide the first simulations that compare these audits beyond a first round and for various round schedules.

Simulations

- We ran simulations to gain further insight into audit behavior and provide additional evidence for theoretical claims.
- We simulated audits for a risk limit of 10% using margins from the 2020 US Presidential election, limiting ourselves to pairwise margins for the two main candidates of 0.05 or larger.
- We used the R2B2 library [?], which provides a framework for the exploration of round-by-round and ballot-by-ballot RLAs. It has implementations of several ballot polling risk-limiting audits as well as a simulator, all written in Python.
- For each of these states, we simulated $10,000 = 10^4$ audits assuming the underlying election was as announced, and an additional $10,000 = 10^4$ audits assuming the underlying election was a tie.

Round Schedules

- A round schedule is a list of positive integers corresponding to the number of ballots sampled in each round. To begin, first round sizes are selected (by search) to achieve a 0.90 probability of stopping assuming the underlying election is as announced.
- Subsequent EoR and SO BRAVO round sizes can be found given the preceding evidence to again achieve a probability of stopping 0.90. It is not known whether MINERVA is risk-limiting for round sizes chosen based on preceding evidence like this, and so subsequent round sizes in MINERVA must be predetermined. We run simulations with MINERVA round schedules where each round has size given by multiplying the previous round size by 1.5 or by 1.0.

Risk

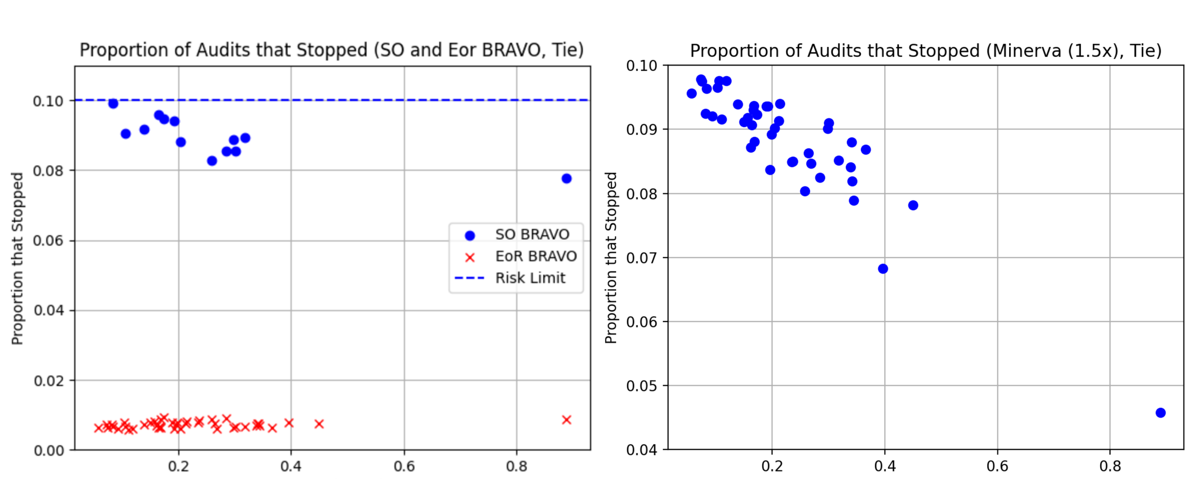

Figure 1. The left hand plot shows the fraction of EoR BRAVO audits (all states with margins at least 0.05) and SO BRAVO audits (the 13 states for which our simulations are complete so far) that stopped in any of the 5 rounds when the underlying election was a tie. The right hand plot, for each state margin, shows the fraction of MINERVA audits with a round size multiple of 1.5 that stopped in any of the 5 rounds when the underlying election was a tie.

Stopping Probability

We are also interested in the probability that an audit will stop in round j given that it did not stop earlier: The conditional stopping probability of an audit $\mathcal{A}$ in round j is

$$\chi_j(\mathcal{A}) = \Pr[\mathcal{A}(X) = Correct\ in\ round\ j \mid H_a \wedge \mathcal{A}(X) \neq Correct\ previously]$$

Experimentally, using our simulations, χ_j would be estimated by the ratio of the audits that stop in round j to those that "entered" round j, i.e. those that did not stop before round j.

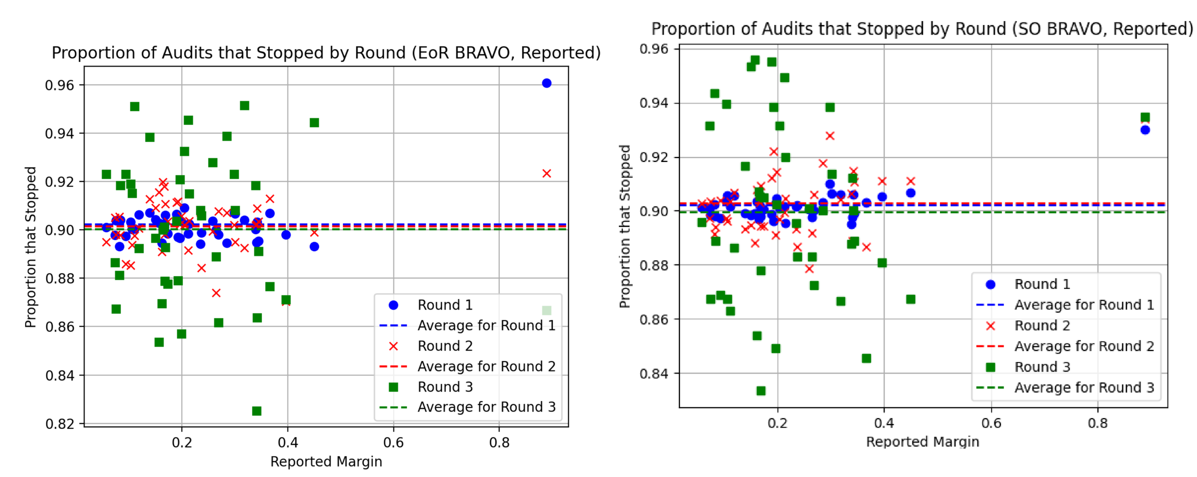

Figure 2. These plots show, for each state margin, when the underlying election is as announced, the number of audits that stopped in the j^{th} round, as a fraction of all audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$ and $\chi_j = 0.9$. The left shows EoR BRAVO and the right shows SO BRAVO.

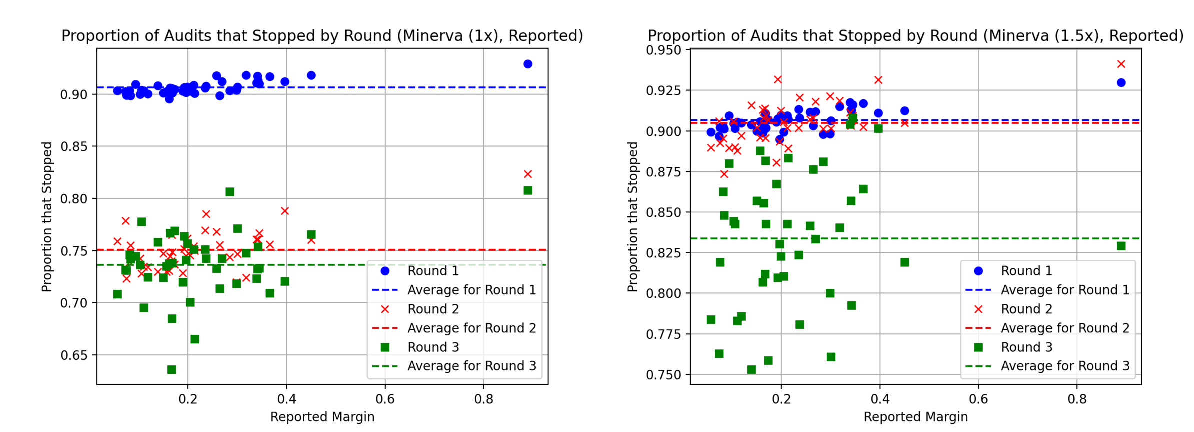

Figure 3. These plots show, for each state margin, when the underlying election is as announced, the number of MINERVA audits that stopped in the j^{th} round, as a fraction of all MINERVA audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$. The left shows MINERVA with the round schedule obtained by $\chi_1 = 0.9$ and round size multiple 1.5.

We can perform a similar study for a first round size with $\chi_1 = 0.25$. See Figure 4 for an example, MINERVA with round mutiplier 1.5.

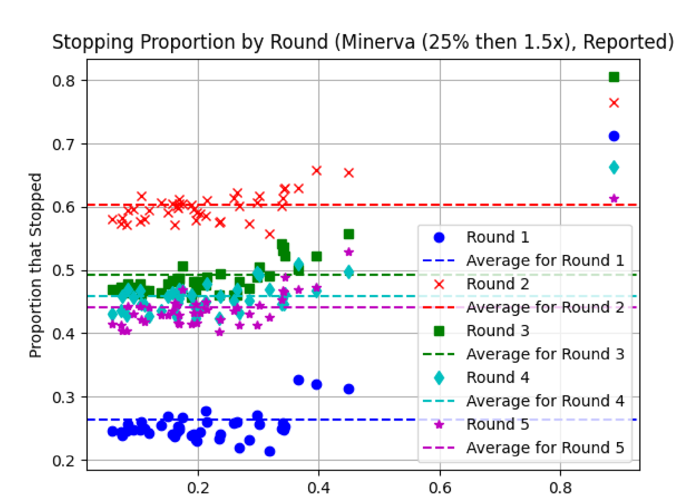

Figure 4. This plot shows, for each state margin, when the underlying election is as announced, the number of MINERVA audits that stopped in the j^{th} round, as a fraction of all MINERVA audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$, round size multiple of 1.5 and $\chi_1 = 0.25$.

Number of Ballots

- The number of ballots sampled is one crude measure of the workload of an audit. To keep the costs of RLAs low, audits should be designed to stop with as few ballots as possible.
- The following plots show the probability of stopping as a function of the average number of ballots sampled by round in our simulations.

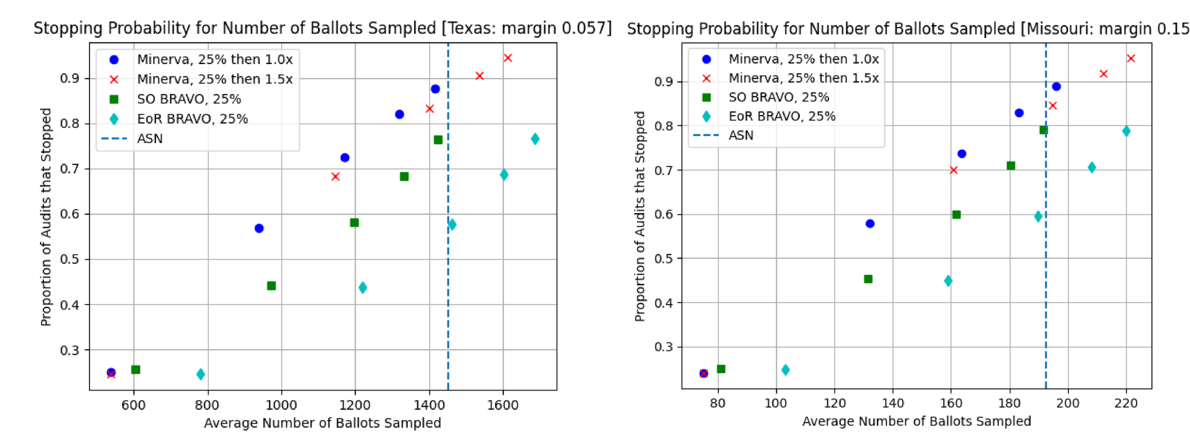

Figure 5. These plots show the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the states of Texas with margin $.057$ (left) and Missouri with margin 0.157 (right), both with first round stopping probability $\chi_1 = 0.25$.

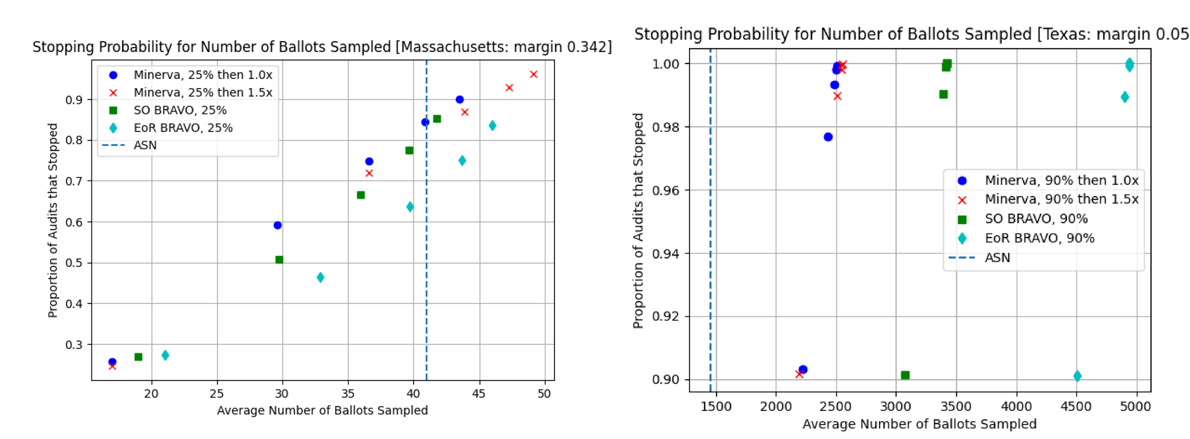

Figure 6. These plots show the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the states Massachusetts (left) with margin 0.342 and $\chi_1 = 0.25$ and Texas (right) with margin 0.057 and $\chi_1 = 0.9$.

For $\chi_1 = 0.25$, the ratio of first round size of EoR BRAVO to MINERVA is 1.45, 1.37, 1.23 for states Texas, Missouri and Massachusetts, and margins 0.057, 0.157 and 0.342 respectively. This may be compared to 2.03, 1.99 and 1.8 respectively for $\chi_1 = 0.9$. Similarly, for $\chi_1 = 0.25$, the ratio of first round size of SO BRAVO to MINERVA is 1.13, 1.08, 1.12 for states Texas, Missouri and Massachusetts, and margins 0.057, 0.157 and 0.342 respectively. This may be compared to 1.38, 1.38 and 1.30 respectively for $\chi_1 = 0.9$.

Modeling Workload

I initially wrote this section since I'd really like to have some evidence that for certain workload parameters, we see an optimal number of rounds greater than 1 with Providence. I am working on that code now.

BRAVO requires the smallest expected number of ballots when ballots are drawn one at a time and the (true) underlying election is as announced. In real audits, election officials draw ballots in rounds because their is some overhead with each round (e.g. opening ballot boxes). Therefore a more sophisticated workload model has a per ballot cost c_b and a per round cost c_r. Let n be the number of ballots sampled and j be the number of rounds. Then the cost C is given by

$$C(n, j) = c_b \cdot n + c_r \cdot j,$$

and depending on the ratio between c_b and c_r, we should expect that certain round schedules achieve lower expected cost.

Providence

- The efficiency of MINERVA is great, but it lacks the flexibility of BRAVO in choosing round sizes based on previous samples.
- PROVIDENCE is our novel RLA which has the efficiency of MINERVA and the flexibility of BRAVO.
- For alternative hypothesis H_a that the election is truly as announced and null hypothesis H_0 that the true election is a tie, BRAVO has the stopping condition that for k cumulative ballots for the winner and n cumulative sampled ballots,

$$\sigma(k, n, p_a, p_0) \triangleq \frac{\Pr[K = k \mid H_a, n]}{\Pr[K = k \mid H_0, n]} \geq \frac{1}{\alpha}.$$

- MINERVA has the stopping condition that in round j with cumulative winner ballots k_j and round sizes $\bar{n}_j = n_1, n_2, \ldots, n_j$

$$\tau_j(k_j, \bar{n}_j, p_a, p_0) \triangleq \frac{\Pr[K_j \geq k_j \wedge \mathcal{A}_{i<j}(X) \neq Correct \mid H_a, \bar{n}_j]}{\Pr[K_j \geq k_j \wedge \mathcal{A}_{i<j}(X) \neq Correct \mid H_0, \bar{n}_j]} \geq \frac{1}{\alpha}.$$

Testing this stopping condition requires computationally expensive convolutions.

- The PROVIDENCE stopping condition uses ideas from both BRAVO and MINERVA and requires no convolution to test:

$$\omega_j(k_{j-1}, k_j, n_{j-1}, n_j, p_a, p_0) \triangleq \sigma(k_{j-1}, n_{j-1}, p_a, p_0) \cdot \tau_1(k_j, n_j, p_a, p_0) \geq \frac{1}{\alpha}.$$

Resistance against an adversary

If adversary $\mathcal{A}$ has knowledge of previous samples and chooses round size n_j in round j, PROVIDENCE is still risk-limiting. Proof of this and the RLA property of PROVIDENCE are currently being revised for submission.

Providence pilot

In February 2022, The Rhode Island Board of Elections hosted a public pilot PROVIDENCE RLA which passed in the first round.

Figure 7. Professor Vora tosses her random-seed-generating 10-sided die, in a roll she dedicated to election officials everywhere.

Simulations

Need to fill in simulation results here. I am thinking (a) risk simulations as evidence that PROVIDENCE is an RLA, and (b) number of ballot results if I get them done quickly enough.

Simulations are currently running to explore the risk, stopping probability, and number of ballots used by PROVIDENCE.

References

Having some latex trouble here, but should include r2b2, the recent simulations paper, bravo paper, athena paper, ...

odbroadrick@gmail.com