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

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 $\textsc{Minerva}$. In a first round with a large sample size (probabilty of stopping at least 90%), $\textsc{Minerva}$ 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 $\textsc{Minerva}$ is risk-limiting for round sizes chosen based on preceding evidence like this, and so subsequent round sizes in $\textsc{Minerva}$ must be predetermined. We run simulations with $\textsc{Minerva}$ round schedules where each round has size given by multiplying the previous round size by 1.5 or by 1.0.

Risk

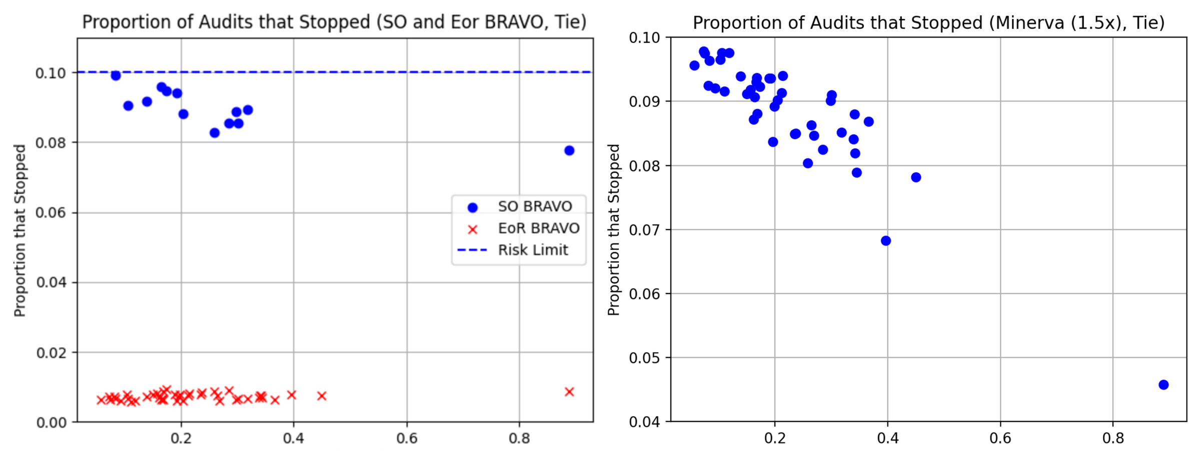

Figure 1. The left hand plot shows the fraction of EoR BRAVO audits (all states with margins at least 0.05) and SO BRAVO audits (the 13 states for which our simulations are complete so far) that stopped in any of the 5 rounds when the underlying election was a tie. The right hand plot, for each state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

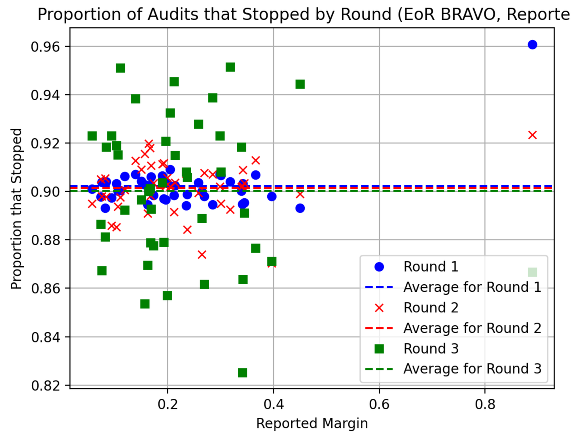

Figure 2. This plot shows, for each state margin, when the underlying election is as announced, the number of EoR BRAVO audits that stopped in the j^{th} round, as a fraction of all EoR BRAVO audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$ and $\chi_j = 0.9$.

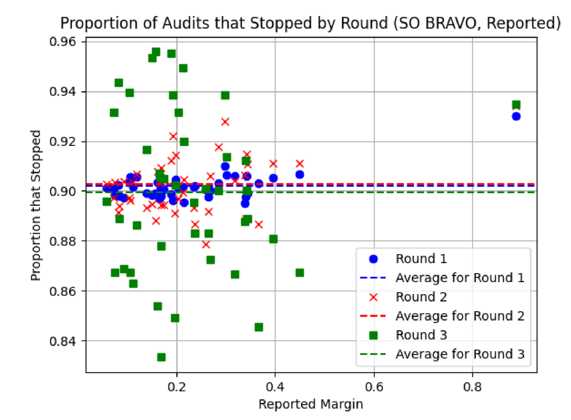

Figure 3. This plot shows, for each state margin, when the underlying election is as announced, the number of SO BRAVO audits that stopped in the j^{th} round, as a fraction of all SO BRAVO audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$ and $\chi_j = 0.9$.

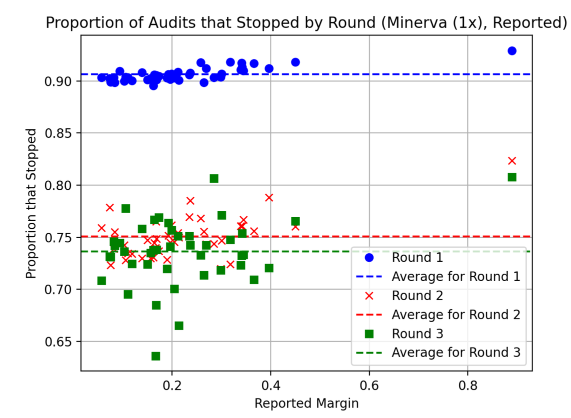

Figure 4. This plot shows, for each state margin, when the underlying election is as announced, the number of $\textsc{Minerva}$ audits that stopped in the j^{th} round, as a fraction of all $\textsc{Minerva}$ audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$, round size multiple of 1.0 and $\chi_1 = 0.9$.

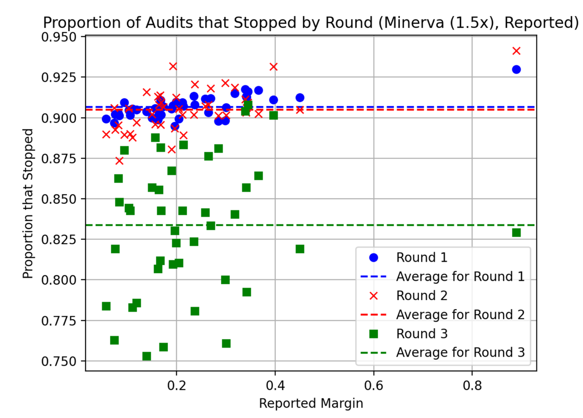

Figure 5. This plot shows, for each state margin, when the underlying election is as announced, the number of $\textsc{Minerva}$ audits that stopped in the j^{th} round, as a fraction of all $\textsc{Minerva}$ audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$, round size multiple of 1.5 and $\chi_1 = 0.9$.

We can perform a similar study for a first round size with $\chi_1 = 0.25$. See Figure 6 for an example, $\textsc{Minerva}$ with round mutiplier 1.5.

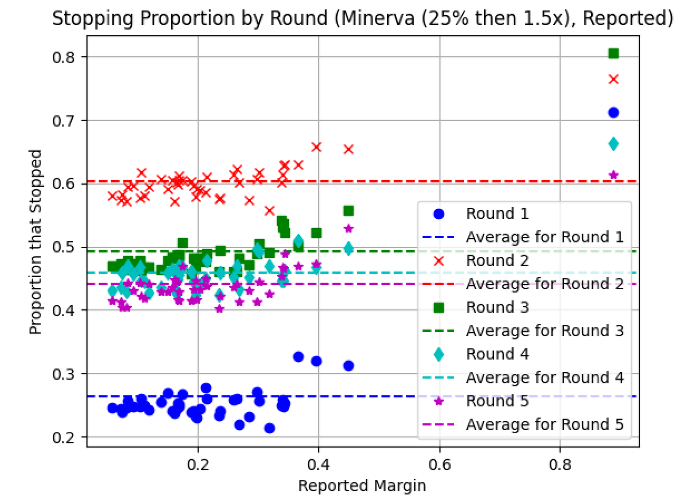

Figure 6. This plot shows, for each state margin, when the underlying election is as announced, the number of $\textsc{Minerva}$ audits that stopped in the j^{th} round, as a fraction of all $\textsc{Minerva}$ audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$, round size multiple of 1.5 and $\chi_1 = 0.25$.

Number of Ballots

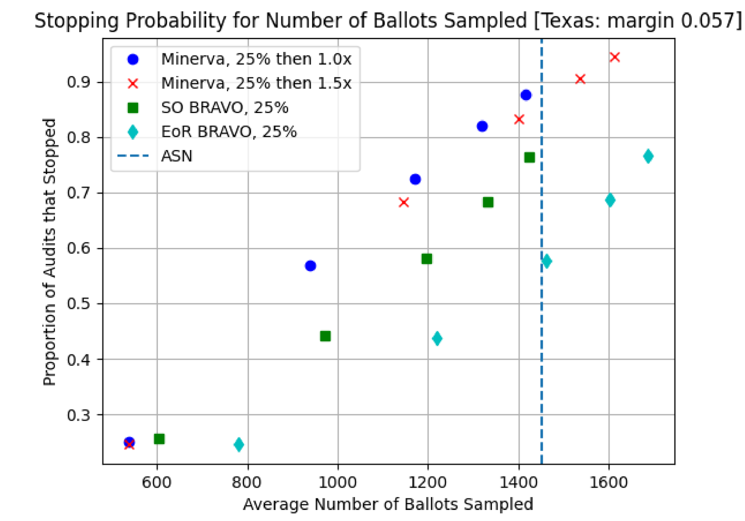

Figure 7. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Texas, margin 0.057, and first round stopping probability $S_1 = 0.25$.

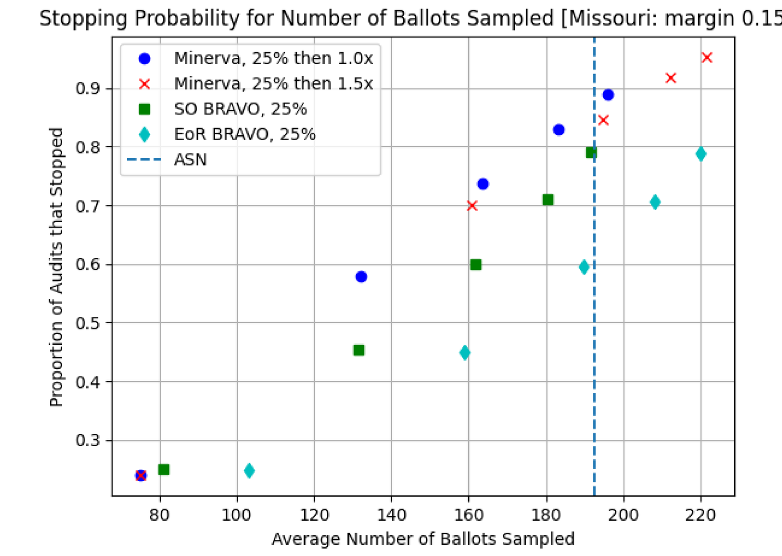

Figure 8. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Missouri, margin 0.157, and first round stopping probability $S_1 = 0.25$.

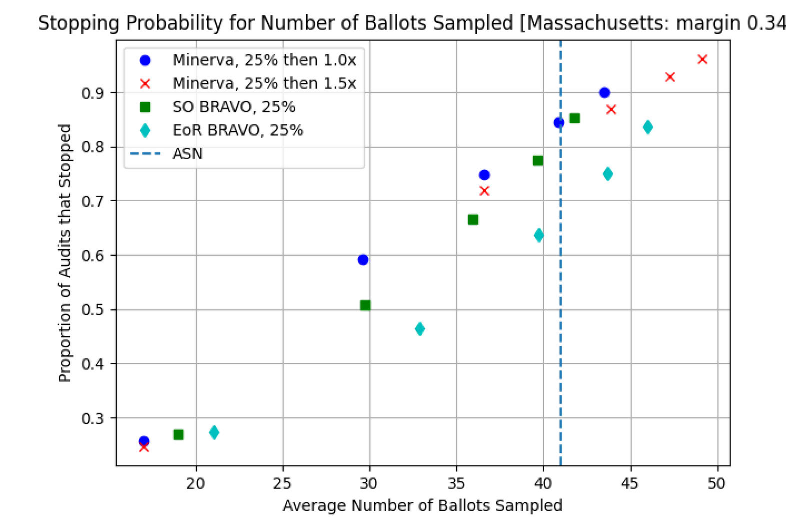

Figure 9. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Massachusetts, margin 0.342, and first round stopping probability $S_1 = 0.25$.

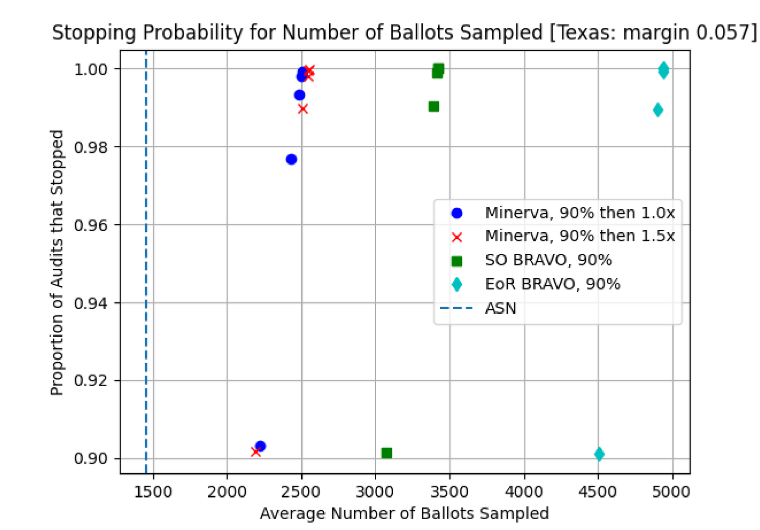

Figure 10. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Texas, margin 0.057, and first round stopping probability $S_1 = 0.9$.

Providence

Providence is great! Efficient like Minerva but flexible like BRAVO.

Simulations

Look, Providence does pretty ok!

References

odbroadrick@gmail.com