PROVIDENCE: an Efficient and Secure Ballot Polling Risk-Limiting Audit

Oliver Broadrick

Thesis Committee:
Arkady Yerukhimovich, Assistant Professor, GW, Committee Chair
Poorvi Vora, Professor, GW, Thesis Advisor, Committee Member
Filip Zagórski, Assistant Professor, University of Wrocław, Thesis Co-Adviser, Committee Member

April 17, 2023

Outline

- Risk-Limiting Audits (RLAs)
- Problem statement and contributions
- Background: From BRAVO, to MINERVA...
- ... to PROVIDENCE
  - Theoretical properties
  - Experimental verification
- Workload models
- Pilot use
- Conclusion

Election security

What do we want?

- The right winner *and* strong evidence that they are the right winner

An approach:

- Cast votes on paper ballots
- Tally the votes with electronic scanners
- Perform compliance and tabulation audits

Risk-Limiting Audit (RLA)

Risk-Limiting Audit (RLA) with risk limit α: A tabulation audit that, if the reported outcome is wrong, will detect and correct it with probability $1 - \alpha$.

α is an error measure. Low values of α are good.

Ballot Polling RLAs

- Assume a voter-verified paper trail stored in numbered boxes, successfully completed compliance audits, and a ballot manifest, which describes how many ballots in each box. Then publicly:

1. Choose a round size, n
2. Select n ballots uniformly at random, with replacement, using a pseudorandom number generator

Ballot Polling RLAs

Rolling the dice!

Ballot Polling RLAs

- Assume a voter-verified paper trail, successfully completed compliance audits, and a ballot manifest. Then publicly:

1. Choose a round size, n
2. Select n ballots uniformly at random, with replacement, using a pseudorandom number generator
3. Find and manually interpret the selected ballots, recording the interpretations
4. Compute the stopping condition $\mathcal{A}$ which outputs:
   (a) Correct: stop, confirming the reported outcome, or
   (b) Undetermined: sample more ballots, or
   (c) Hand count: stop and perform a full recount
5. If more ballots are to be drawn, return to step 1 (or, when election officials choose, proceed to a full, manual recount)

Existing ballot polling RLAs

BRAVO

- In the two candidate case is an instance of Wald's classic Sequential Probability Ratio Test (SPRT)
- Most efficient RLA when the stopping condition is checked after each ballot is drawn (ballot-by-ballot)
- In real audits, decisions are taken after many ballots are drawn (round-by-round), and BRAVO is implemented as:
  - Selection-Ordered (SO) BRAVO, where ballot selection order is retained, and the decisions are taken as though the audit were ballot-by-ballot
  - End-of-Round (EoR) BRAVO, where the decision using the BRAVO stopping rule is taken once, after the entire round of ballots is drawn

Existing ballot polling RLAs

MINERVA

- Recent RLA designed for round-by-round use
- Known to be risk-limiting if all round sizes are predetermined, before the audit begins
- In a first round chosen to give a 0.90 probability of stopping, MINERVA requires
  - 50% as many ballots as EoR BRAVO
  - 70-80% as many ballots as SO BRAVO
- For smaller stopping probability rounds, the benefit of MINERVA decreases (i.e. as the audit approaches the ballot-by-ballot case)

Problems We Will Address

1. Predetermined round sizes are limiting
    - e.g. After nearly meeting the stopping condition in a round, MINERVA requires escalation to the predetermined next round size
    - May be more efficient to choose future round sizes as a function of previous samples; would allow greater flexibility
2. Existing workload measures don't capture the cost of a round
    - We are unaware of any RLAs that have ever actually drawn a single ballot at a time
        - More efficient to draw many at once

Contributions

1. RLA PROVIDENCE
   - An audit with the efficiency of MINERVA and flexibility of BRAVO
   - Proofs of theoretical properties, plus experiments
2. Evaluation of BRAVO, MINERVA, and PROVIDENCE under simple, new workload models that account for the cost of a round

Background

RLA definition

An audit $\mathcal{A}$ is a function from the space of possible samples, $\mathcal{X}$, to the set $\{\text{Correct, Undetermined}\}$ where:
(a) Correct: stop the audit
(b) Undetermined: sample more ballots

Definition (α-Risk-Limiting Audit (α-RLA))

An audit $\mathcal{A}$ is an α-RLA if for all samples $X \in \mathcal{X}$

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

where H_0 corresponds to the incorrectly reported outcome *closest* to the reported outcome (i.e. a tie)

Binary hypothesis test:

- H_a, the outcome was correctly reported
  - $p = p_a$ the reported proportion
- H_0, the outcome was incorrectly reported
  - $p = \frac{1}{2}$ the reported proportion

Risk $:= \Pr[\mathcal{A}(X) = \textit{Correct} \mid H_0]$
Stopping Probability $:= \Pr[\mathcal{A}(X) = \textit{Correct} \mid H_a]$

Notation:
n_j is the cumulative round size in round j
k_j is the cumulative tally of winner ballots in round j

Definitions

Bravo:

$$\sigma_j(k_j) \triangleq \frac{p_a^{k_j}(1-p_a)^{n_j-k_j}}{p_0^{k}(1-p_0)^{n_j-k_j}} = \frac{\Pr[K_j = k_j | H_a]}{\Pr[K_j = k_j | H_0]} \geq \frac{1}{\alpha}$$

Minerva:

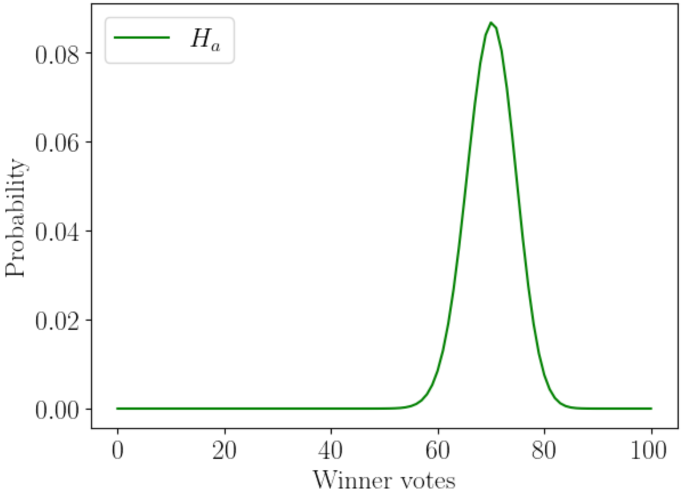
$$\tau_j(k_j) \triangleq \frac{Pr[K_j \geq k_j \wedge \forall_{i<j}(\mathcal{A}(X_i) \ \neq \text{Correct}) \mid H_a, \mathbf{n_j}]}{Pr[K_j \geq k_j \wedge \forall_{i<j}(\mathcal{A}(X_i) \ \neq \text{Correct}) \mid H_0, \mathbf{n_j}]} \geq \frac{1}{\alpha}$$

from BRAVO to MINERVA

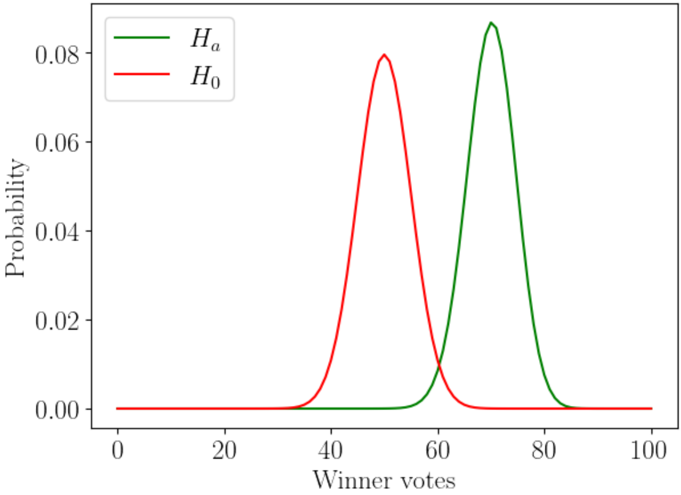
A toy example:

- Two candidates
- Reported proportion of votes for the winner $p_a = 0.7$
- Risk limit $\alpha = 0.1$
- First round size $n_1 = 100$

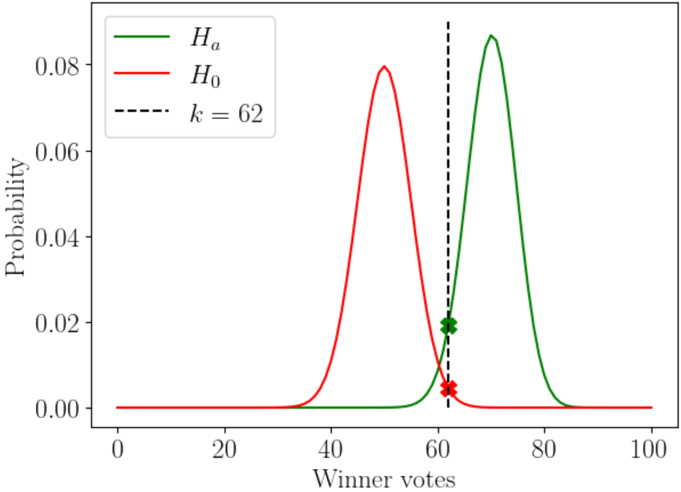
from Bravo to Minerva



from BRAVO to MINERVA

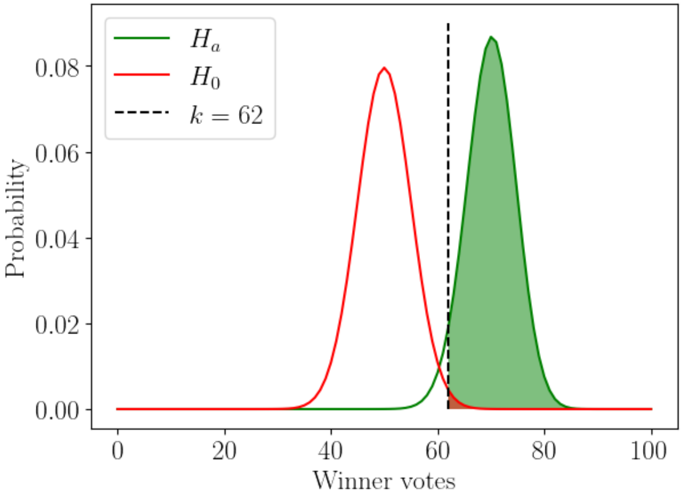


from BRAVO to MINERVA



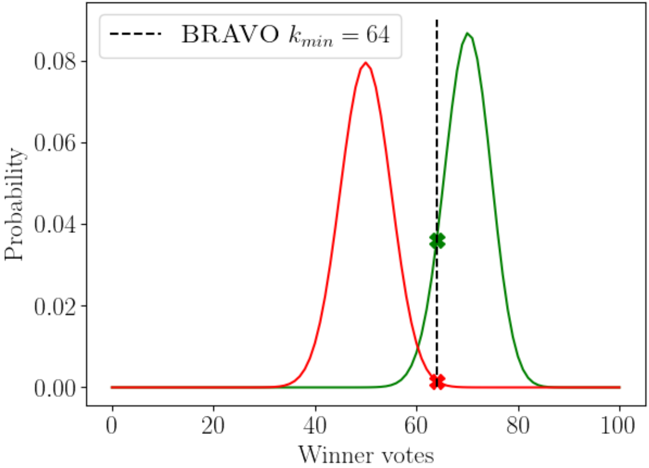
$$\text{BRAVO: } \frac{\Pr[K_j = k_j | H_a]}{\Pr[K_j = k_j | H_0]} = \sigma(62) \approx 4.26 < 10 = \frac{1}{\alpha}$$

from Bravo to Minerva



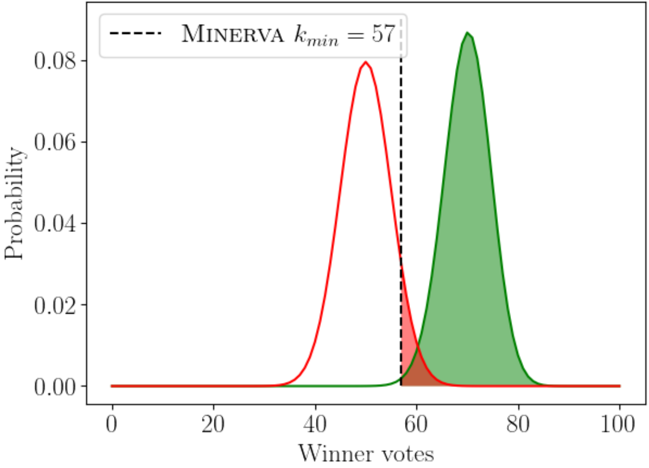
$$\textsc{Minerva}: \ \frac{Pr[K_j \geq k_j \mid H_a, \mathbf{n_j}]}{Pr[K_j \geq k_j \mid H_0, \mathbf{n_j}]} = \tau(62) \approx 92.10 \geq 10 = \frac{1}{\alpha}$$

from BRAVO to MINERVA



$$\sigma(63) \approx 9.95 < 10 \leq 23.21 \approx \sigma(64)$$

from BRAVO to MINERVA



$$\tau(56) \approx 7.37 < 10 \leq 10.32 \approx \tau(57)$$

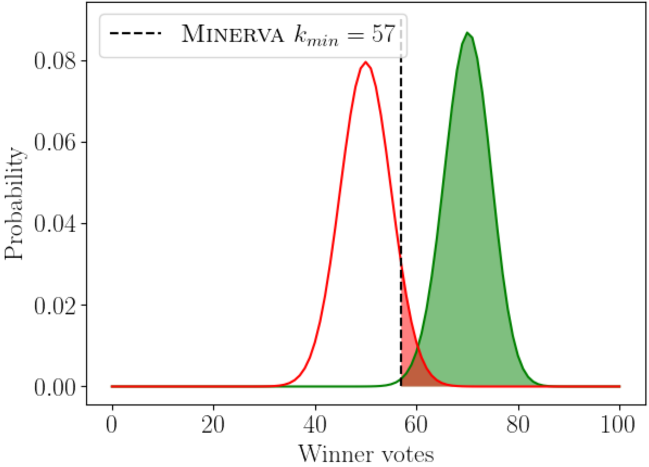
Why does this tail ratio give an RLA?

Let $R_j = \Pr[\,\text{stop in round j and no earlier} \mid H_0\,]$
and $S_j = \Pr[\,\text{stop in round j and no earlier} \mid H_a\,]$.
Then,

$$\frac{S_j}{R_j} \geq \frac{1}{\alpha} \implies R_j \leq \alpha S_j \implies \sum_j R_j \leq \alpha \sum_j S_j \leq \alpha$$

So, audits that enforce $\left(\dfrac{S_j}{R_j} \geq \dfrac{1}{\alpha}\right) \forall j$ are risk-limiting!

from BRAVO to MINERVA

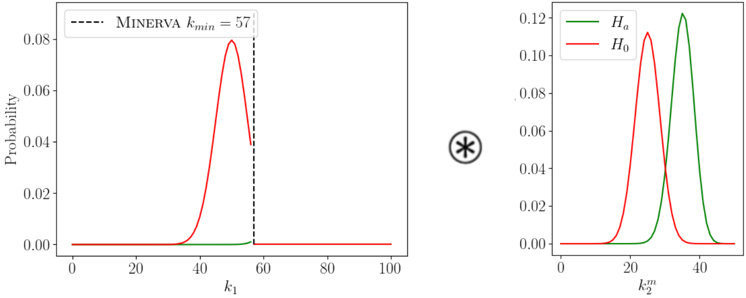


$$\tau(56) \approx 7.37 < 10 \leq 10.32 \approx \tau(57)$$

from BRAVO to MINERVA

$$\tau_j(k_j) \triangleq \frac{Pr[K_j \geq k_j \wedge \forall_{i<j}(\mathcal{A}(X_i) \neq \text{Correct}) \mid H_a, \mathbf{n_j}]}{Pr[K_j \geq k_j \wedge \forall_{i<j}(\mathcal{A}(X_i) \neq \text{Correct}) \mid H_0, \mathbf{n_j}]} \geq \frac{1}{\alpha}$$

If $n_2 = 150$, what's pdf for $K_2 = k_2 \wedge \mathcal{A}(X_1) \neq Correct$?
What's $\Pr[K_2 = 100 \wedge \mathcal{A}(X_1) \neq Correct]$?



Implicitly assumes that n_2 is the same for all k_1

Our work

Adversarial round sizes

Goal of adversary: increase the risk above α

Definition (Weakly round-choosing adversary)

One who, before the audit begins, selects round sizes n_j for all j as a function of audit and contest parameters:

$$n_j(\alpha, p_a, p_0, ballot\_manifest)$$

EoR Bravo, SO Bravo, Minerva are all resistant to a weakly round-choosing adversary

Adversarial round sizes

Definition (Strongly Round-Choosing Adversary)

One who selects n_1:

$$n_1(\alpha, p_a, p_0, \mathit{ballot\_manifest}),$$

and n_j for $j \geq 2$:

$$n_j(\alpha, p_a, p_0, \mathit{ballot\_manifest}, \mathbf{k_{j-1}}, \mathbf{n_{j-1}})$$

EoR Bravo, SO Bravo are resistant to a strongly round-choosing adversary; Minerva is not

Providence

Providence:

$$\omega_j \triangleq \frac{Pr[K_j \geq k_j \wedge K_{j-1} = k_{j-1} \mid H_a, n_{j-1}, n_j]}{Pr[K_j \geq k_j \wedge K_{j-1} = k_{j-1} \mid H_0, n_{j-1}, n_j]} \geq \frac{1}{\alpha}$$

RLA proof idea:

- Sum over all k_{j-1}
- Again, $\left(\frac{S_j}{R_j} \geq \frac{1}{\alpha}\right) \forall j \implies$ RLA

$$\omega_j = \frac{\Pr[K_{j-1} = k_{j-1} \mid H_a, n_{j-1}] Pr[K_j^m \geq k_j^m \mid k_{j-1}, H_a, n_{j-1}, n_j]}{\Pr[K_{j-1} = k_{j-1} \mid H_0, n_{j-1}] Pr[K_j^m \geq k_j^m \mid k_{j-1}, H_0, n_{j-1}, n_j]}$$

$$\omega_j(k_j, k_{j-1}) = \sigma(k_{j-1})\tau_1(k_j^m)$$

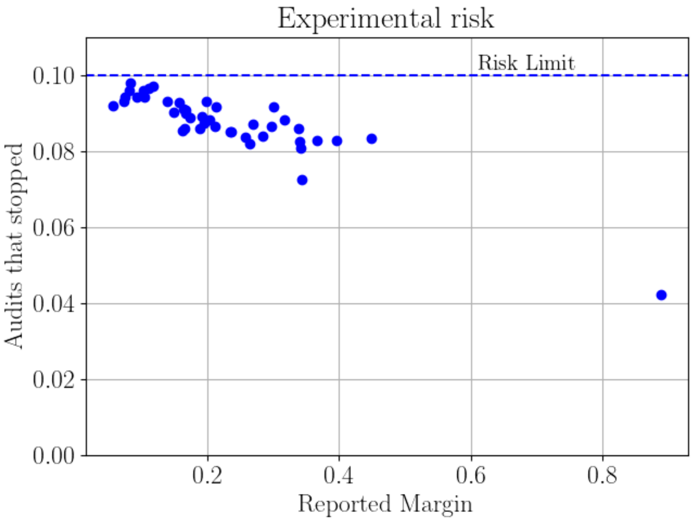
Simulations

What if Oliver put some errors in the proof?

Simulations:

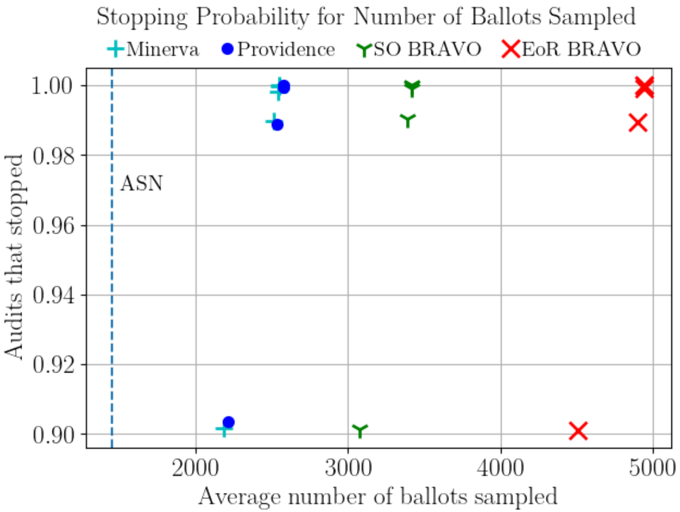
- 2020 Presidential Contest (states with a pairwise margin at least 0.05)
- Risk limit $\alpha = 0.1$
- 10^4 trials per state assuming H_a (i.e. p as reported)
- 10^4 trials per state assuming H_0 (i.e. $p = 0.5$)
- 5 rounds
- Round sizes chosen to each give stopping probability of 0.9, conditioned on the current sample (except Minerva which uses first round size to achieve 0.9 stopping probability and then a multiplier of 1.5)

PROVIDENCE risk



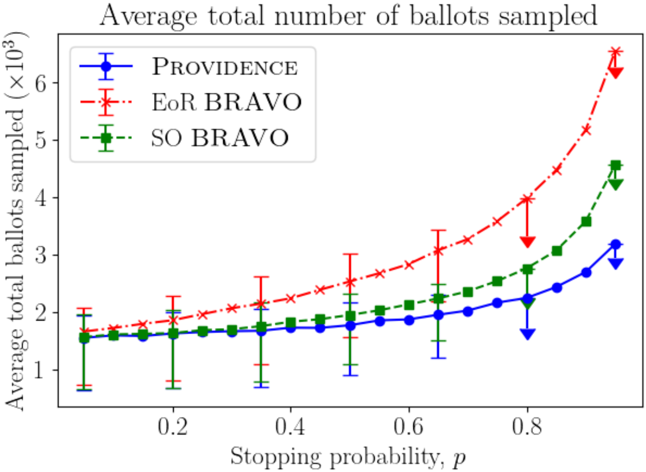
Number of ballots

Texas, with margin 0.057



Number of ballots

Now parameterize round sizes by stopping probability
2016 Presidential contest in VA with margin ≈ 0.053



Workload

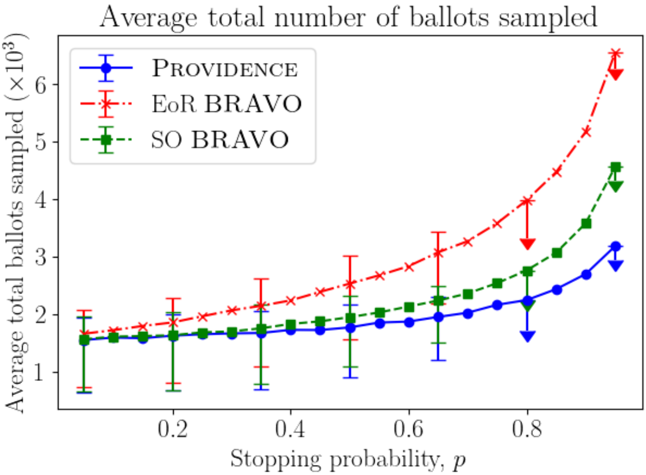
With a round cost:

$$W(E_b, E_r) = E_b c_b + E_r c_r$$

E_b: expected number of ballots
E_r: expected number of rounds
c_b: per ballot cost
c_r: per round cost

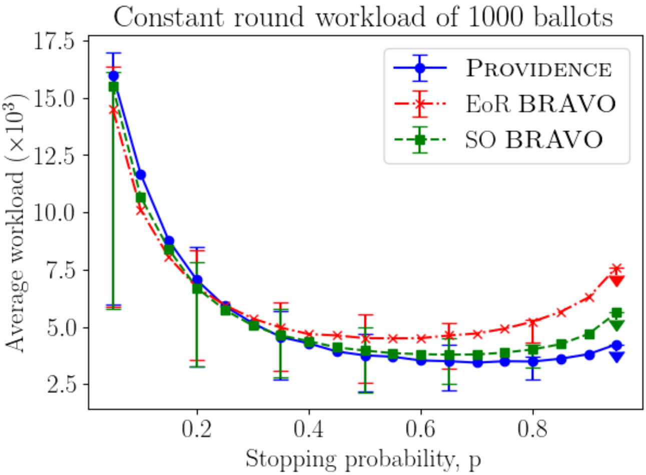
Number of ballots

$$W(E_b, E_r) = E_b c_b + E_r c_r \text{ with } c_b = 1 \text{ and } c_r = 0$$

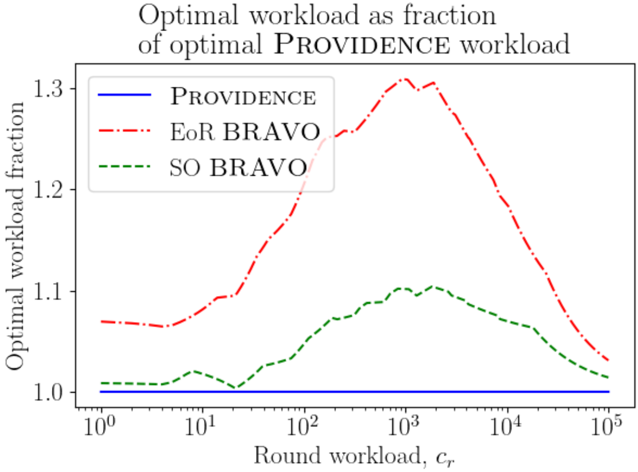


Workload

$W(E_b, E_r) = E_b c_b + E_r c_r$ with $c_b = 1$ and $c_r = 1000$



Optimal workload



Pilot audit

February 2022 in Providence, Rhode Island

Margin ≈ 0.256

ballots	Providence	Minerva	SO Bravo	EoR Bravo
140	**0.0418**	**0.0418**	**0.0541**	0.366

Conclusion

- PROVIDENCE: efficient and secure
- Piloted in RI, and implemented in Arlo, most popular RLA software
- Introduction of workload models accounting for round and precinct costs; misleading samples
- Future work: Optimal audits
  - BRAVO is SPRT which is optimal for the round schedule of $[1, 1, 1, \ldots]$
  - MINERVA and PROVIDENCE simplify to BRAVO for this case
  - Nothing is known about the existence of an optimal test for arbitrary round schedules $[n_1, n_2, n_3, \ldots]$
  - Application of round-by-round tests to other domains (e.g. medical trials)

Thank you for listening