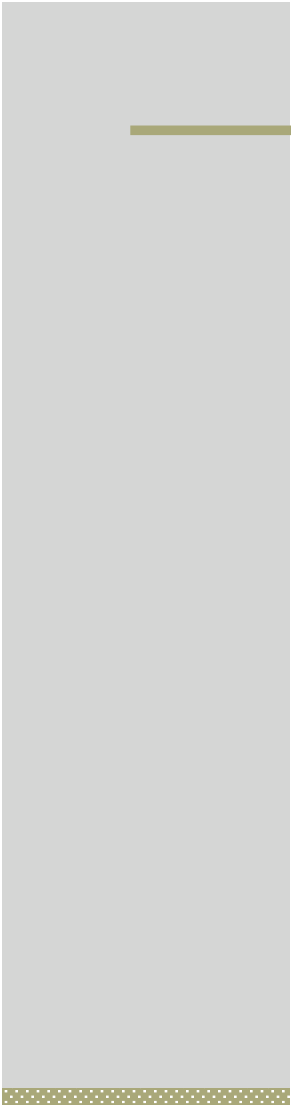




# On Auditing Elections When Precincts Have Different Sizes



**Javed A. Aslam**  
College of Computer and  
Information Science  
Northeastern University




**Raluca A. Popa and Ronald L. Rivest**  
Computer Science and Artificial  
Intelligence Laboratory  
M.I.T.

# Outline

---

- Auditing Overview
- Motivation
- Methods
  - NegExp
  - PPEBWR
- Evaluation
- Recommendations
- Conclusions

# What Is Auditing?

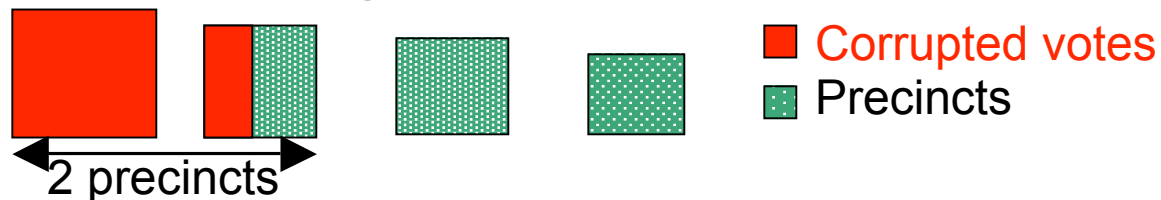
- Post-election auditing is useful for detecting accidental or malicious errors
- Precinct auditing procedure:
  - Determine the set of precincts to audit
    - Use randomization 
  - Hand count paper ballots in sampled precincts
  - Compare hand count to electronic tally:
    - If sufficiently close, declare electronic result final 
    - If significantly different, investigate! 

# How to Select Precincts?

1. Fixed audit
  - Fixed number or percentage of precincts
  - Shown to be insufficiently accurate or inefficient
2. Margin-dependent audit
  - Based on **margin of victory** (winner votes – runner-up votes)
  - Half margin of victory is *least number of corrupted votes*
  - Achieves a desired level of confidence
  - Typically precincts sampled with equal probability
3. Size and margin dependent audit
  - Sample with probabilities dependent on **precinct sizes**
  - **Provides substantial savings!**

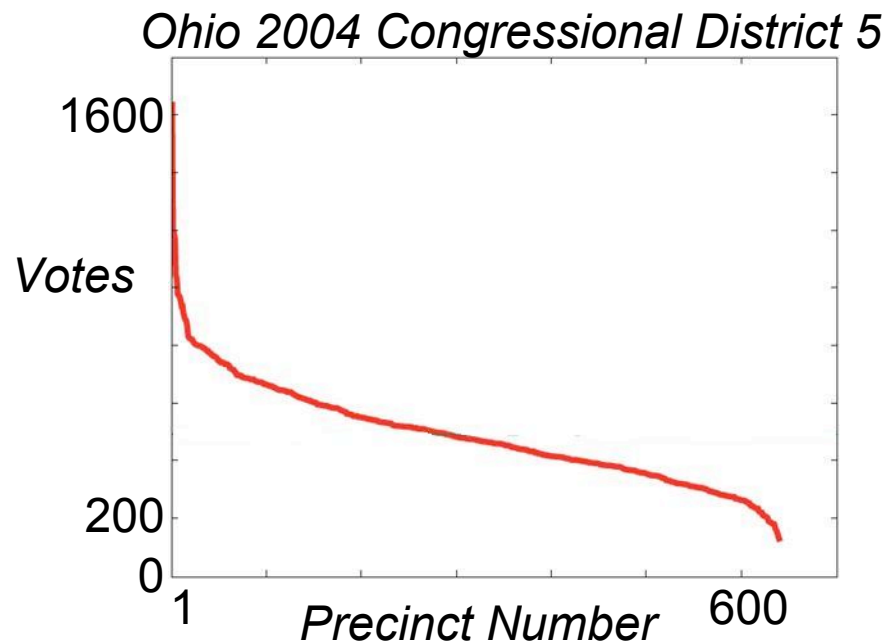
# Previous Work

- *SAFE* [McCarthy et al., 2007]
  - Compute least number of corrupted votes from margin of victory
  - Compute least number of corrupted precincts
    - Assume larger precincts are corrupted first



- Precincts are audited with *equal* probability
- Sample size ensures desired level of confidence
- Inefficient when precinct sizes vary significantly
- **Our methods reduce the workload by about half**

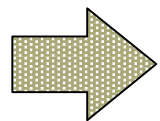
# Motivation



- Precinct sizes vary greatly
  - Largest: 1637 votes
  - Smallest: 132 votes
  - **More than an order of magnitude!**
- Larger precincts can allow greater fraud
  - Should audit with higher probability

# Goal

---

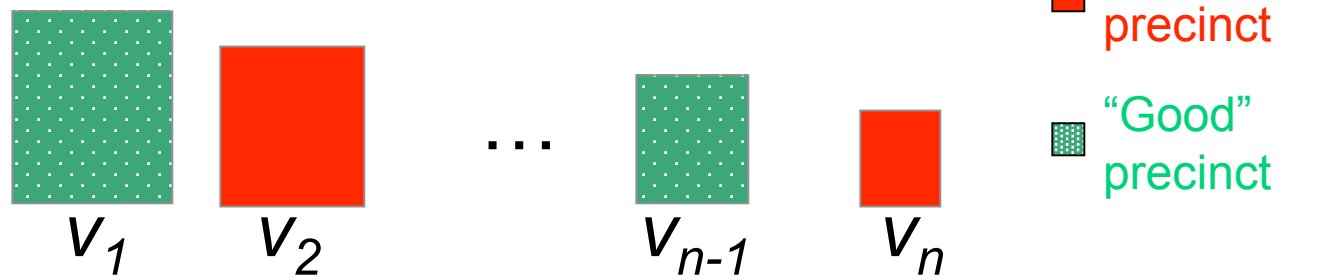


Devise efficient auditing procedures by considering precinct sizes

- Significance (confidence):
  - If the election result is corrupted, at least one corrupted precinct is detected at the desired significance
  - If no fraud is detected, the election result is certified at the desired significance
- Efficiency: \$
  - Few votes and precincts audited

# Model

- Example: Ohio 2004 Congressional District 5
- $n$  precincts
  - $n = 640$  precincts



- $v_i =$  number of votes in precinct  $i$ 
  - $v_1 \dots v_n = 1637 \dots 132$  votes
- $V =$  total number of votes ( $\sum v_i$ )
  - $V = 315,540$  votes



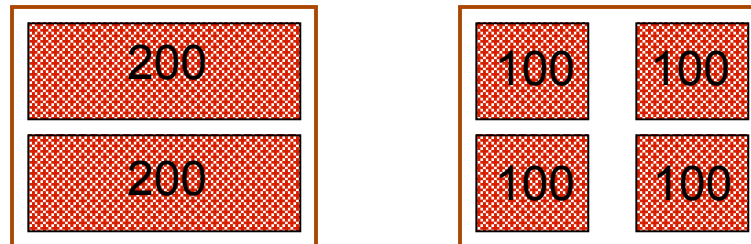
# Model (cont'd)

- $M$  = margin of victory in votes
  - Vote difference between winner and runner-up
  - $M/2$  is least number of corrupted votes if election is fraudulent
  - If winner won by 1% over the runner-up,  $M = 3,155$  votes
- $\alpha$  = desired significance level
  - 1 - confidence level
  - 8% (confidence of 92%)

# Approach

Each precinct is audited with a probability dependent on its size,  $v_j$ .

- Sets of same total size have about the same probability of being audited:



- Paper presents error bounds instead of sizes
  - $kv_j$ ,  $k = 0.4$  [Dopp and Stenger, 2006]

# Our Methods

---

- Two methods:
  - *NegExp*
    - Each precinct is audited independently with a probability dependent on its size
  - *PPEBWR*
    - One precinct is selected during each of a sequence of rounds with a probability proportional to its size
- Both ensure the desired significance level independent of the adversarial strategy

# NegExp Method

- “Negative Exponential”
- Audit each precinct independently with probability:

$$p_i = 1 - \beta^{-v_i}, \quad \beta = \text{constant}$$

- The chance of auditing at least one precinct from a set of precincts is given by the **total size**

- Example: a set of two precincts  $i$  and  $j$

$$1 - (1 - p_i)(1 - p_j) = 1 - \beta^{-v_i} \beta^{-v_j} = 1 - \beta^{-(v_i + v_j)}$$

- Condition for significance level:

$$\beta = \alpha^{-k/M} \Rightarrow p_i = 1 - \alpha^{v_i \cdot k/M}$$

# PPEBWR Method

- “Probability proportional to error bound (size) with replacement”

- During each round, one precinct is selected with the probability distribution:

$$(v_1/V, \dots, v_n/V)$$

- Repetitions (rare) audited only once
- Number of rounds for the desired significance:

$$\frac{\ln \alpha}{\ln(1 - M/kV)}$$

# Example

- Largest precinct:  $v_1 = 1637$  votes
- Smallest precinct:  $v_n = 132$  votes
- NegExp:
  - $p_1 = 41\%$ ,  $p_n = 4.1\%$
- PPEBWR:
  - During each round:  $p_1 = 0.52\%$ ,  $p_n = 0.042\%$
  - Over all the rounds:  $p_1 = 40\%$ ,  $p_n = 4.1\%$
- Both have similar final auditing probabilities

# Dice Rolls in NegExp

- Audit a precinct with probability  $p$ :
  - Roll four ten-sided dice to get a four-decimal number



- Audit the precinct if the result is smaller than  $p$
- Example:
  - $p_1 = 0.41 \rightarrow$  audit
  - $p_n = 0.041 \rightarrow$  do not audit

# Dice Rolls in PPEBWR

- Audit a precinct from the distribution:

$$(v_1/V, \dots, v_n/V)$$

- Consider each vote labeled from 1 to  $V$  and select a vote number at random
  - Roll a ten-sided die for each digit



- Repeat until number is from 1 to  $V$
- Audit the precinct containing the vote



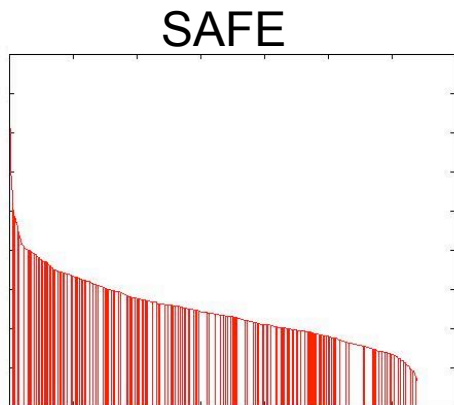
# Comparison to SAFE

---

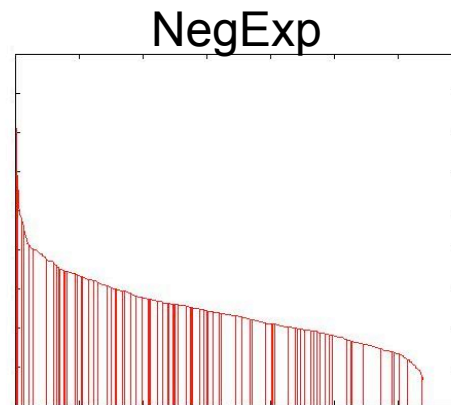
- Ohio 2004 Congressional District 5
- $\_ = 8\%$
- Margin of victory 1%
- Expected number of votes to audit ( $\sum v_i p_i$ )
  - SAFE: 95,155 (30%)
  - NegExp: 50,937 (16%)
  - PPEBWR: 50,402 (16%)

# Comparison to SAFE (cont'd)

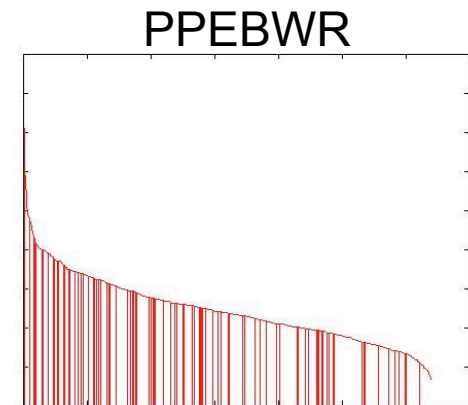
- Expected number of precincts audited ( $\sum p_i$ )
  - Votes versus precinct number for **audited precincts**:



193 precincts (30%)



Mean: 92.6 precincts  
(14%)

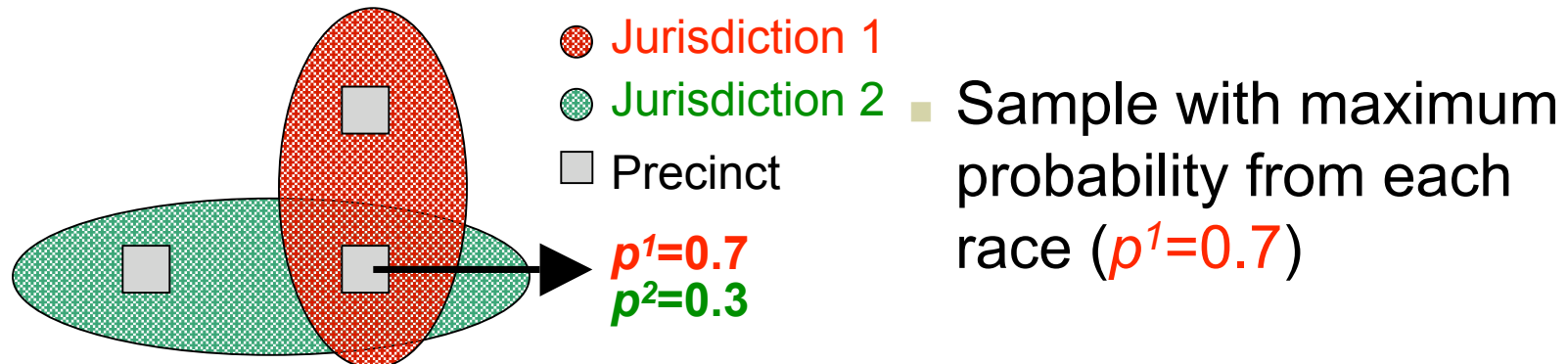


Mean: 91.6 precincts  
(14%)

- **About twice as efficient**

# NegExp vs. PPEBWR

- NegExp is more *flexible*:
  - Races with overlapping jurisdictions




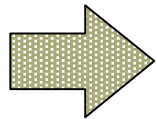
- Adjusting auditing probabilities
  - Remember dice roll outcome and decide whether to audit or not

➔ Recommended where flexibility is needed

# NegExp vs. PPEBWR (cont'd)

---

- PPEBWR is more *efficient*
  - Slightly less precincts and votes audited on average
  - Less dice rolls 
    - NegExp rolls dice per precinct (eg. 640)
    - PPEBWR rolls dice per round (eg. 100)



Recommended for simple elections

# Conclusions

---

- Two new practical auditing procedures based on precinct sizes
  - NegExp
  - PPEBWR
- About twice as efficient as previous approaches

Thank you!