E2E to Hand-to-Eye
Verifiability, Trust, Audits

Philip B. Stark

Department of Statistics
University of California, Berkeley

VoteID 13
University of Surrey
Guildford, England
17–19 July 2013

Credit—and blame—where due

Grateful to J. Benaloh, O. Pereira, R. Rivest, P. Ryan, V. Teague, P. Vora.
I'm ignorant of much of which I shall speak. It's not their fault.

Research is when you don't know what you're doing

Lots of "research" in this talk:
I don't know what I'm talking about, nor what I want to say.

Statistics: Trust, but Waffle

Stereotypes of two communities

Gross generalizations about CS folks:

- More attention to tamper resistance and tamper evidence than to resilience
- Emphasize "hardening" systems to prevent problems
- More focus on detecting problems than on correcting problems
- Assume adversaries are malicious, but allies behave randomly (in a helpful way)

Gross generalizations about Stat folks:

- Assume there will be a problem
- Emphasize estimating the size of the problem
- More focus on adapting to perturbations
- Assume adversaries behave randomly, but question whether allies do (random $\neq$ haphazard)

The difference between an optimist and a pessimist is . . .

I think:

- There will be a problem
- Emphasize correcting the problem if possible
- Adversary might be malicious
- Allies not random, just erratic, uncompliant, befuddled, lazy, malicious (cf J.P. Clark's law)
- Good to harden systems, but: diminishing returns, high monetary and complexity costs, perfection impossible
- Want system that is resilient even against malicious adversaries and erratic allies and reports failures

Wallach's Insight (D. Wallach)

The purpose of an election is to convince the loser s/he lost.

Evidence-Based Elections (PBS, D. Wagner)

Elections officials should provide convincing evidence that the outcomes are right, or admit that no such evidence is forthcoming.

What's Convincing?

- Depends on whom/what you trust—and for what.
- One person's "obviously!" is another's "seriously?"
- Is there a reasonable standard of "reasonable person"?

which brings us to Unicorns

'Unicorn': noun.

- Something that in principle could exist.
- But there's no convincing evidence it does.
  (Recursion unintentional)

Some election-related Unicorns

- universal suffrage
- perfect registration system
- adequate provisioning of election-day supplies
- perfectly usable ballot; perfectly marked ballots
- perfectly attentive or compliant voter
- perfectly attentive or compliant pollworker
- perfectly reliable hardware
- perfectly secure server, client, or PBB
- perfect physical security of hardware or ballots
- unconditional software independence
- perfect audit trail
- exhaustive list of attacks or attacker capability
- a second “reasonable person”

Conceptual and Technical Tools

- Software Independence and Strong Software Independence
- Risk-limiting audits (and random sampling in general)
- Compliance audits
- Resilient canvass frameworks
- End-to-End verifiability
- Public bulletin boards (PBBs)
- Cryptographic commitments
- Homomorphic encryption
- Mixnets
- Cut-and-choose and the “Benaloh challenge”
- Zero-knowledge proofs (ZKP and NIZKP)

[Strong] Software Independence (Rivest & Wack)

Undetected change or error in software cannot produce an undetectable change or error in the results [and can reconstruct the correct result without re-running the election].

- Property of election, not equipment
- System can produce wonderful voter-verified paper trail and still not be SI, if paper trail is not curated adequately
- SI guarantees that you can tell whether something went wrong, but not that anyone will bother to check
- SSI guarantees that the right outcome could be found without re-running the election, but you still gotta look and do the work

Risk-limiting Audit (PBS)

Known minimum probability that the audit will correct the outcome if the outcome is wrong, no matter why the outcome is wrong.

- Property of audit: isn't a particular procedure
- Requires SSI voting system: adequately accurate audit trail
- Typical strategy: H2E inspection of ballots until either there's strong evidence that the outcome is right—or until all ballots have been counted by hand, revealing correct outcome
- Generates quantitative evidence

Compliance Audit

Seek convincing affirmative evidence that audit trail reflects correct outcome.

- Checks whether system, as deployed, is SSI on audit day
- Ballot accounting, physical security checks, chain of custody checks, etc.
- Generates qualitative evidence

Resilient Canvass Framework (Benaloh, Jones, Lazarus, Lindeman, PBS)

Known minimum chance that if the overall canvass (human, procedural, & machine elements) declares an outcome, that outcome is correct.

- System should be self-correcting or admit that the "perturbation" may have exceeded its fault tolerance
- Property of election, not just equipment
- Combines potentially SSI system with compliance audit and RLA
- If compliance audit doesn't find convincing evidence that system—as deployed—was SSI, abort; else, perform RLA
- Combines qualitative and quantitative evidence

E2E

Personal verifiability: voter can verify whether her vote was cast as intended and included in the tally.
Universal verifiability: anyone can verify whether the published votes were tabulated correctly.

- Property of election, not equipment.
- Is it enough to ask?
- Is it more than necessary?

Claim: Much Ado about orthogonal Issues

Cast as intended, recorded as cast, counted as recorded, reported as counted, . . .
Nice, but not entirely relevant to whether outcome is right.

Want trustworthy Outcomes

Example of verifying outcome w/o verifying tabulation: ballot-polling audit (more later).

Evidence-based Elections (PBS & Wagner)

Evidence = Auditability + Auditing

- LEOs should provide convincing evidence that the outcome is right
- current elections in US and elsewhere procedure-based: equipment certification and election process
- EBE puts incentives in the right place: improving transparency, procedures, equipment, curation, etc., means less work for LEOs to generate convincing evidence

What/whom do elections require us to trust?

Varies widely. Trust for accuracy differs from trust for anonymity.
Might include:

- ourselves
- other voters, “helper organizations”
- vendors of hardware and services
- hardware, hardware designers, hardware manufacturers
- software and programmers
- elections officials
- pollworkers
- cryptography & cryptographers, ZKP, NIZKP
- statistics & statisticians, randomness, dice, PRNGs
- physical and information security measures

I confess . . .

- There are voters I'd trust more than election officials, and vice versa.
- There are cryptographers I'd trust more than statisticians, and vice versa.
- There are pollworkers I'd trust more than vendors. (Not sure about the converse.)
- I don't know how much I'd trust helper organizations. Nor whether they exist. (Unicorn?)

Trust whom, for what?

- Evidence about outcomes? Assured anonymity? Public confidence?
- How hard is it for the trusted party to do her job?
- What is the consequence of failure?
- Is the trustee the potential attacker? If not, who is?
- How easy is it to discover failures?
- What failures can be recovered from?
- How hard/expensive/slow is it?
- How can current systems be augmented to improve resilience?

Proposed notional goals for voting systems

1. give convincing evidence that outcome is right, or fess up
2. be affordable, practical, maintainable, explainable
3. robustify more than harden: less brittle, more resilient
4. minimize reliance on unicorns
5. parallelize trust requirements

“Parallelizing” the trust requirement

Would like choice in whom to trust and to eliminate single points of failure.

For various components of the election, can be done with:

- transparency (plus trusted observers): publishing code, algorithms, etc.; allowing tally and audit observers; webcams
- threshold encryption
- allowing observers to contribute to PRNG seed for audits
- E2E combined with hand-to-eye audits of paper

Audits

What do we want election audits to do?

- Ensure that the electoral outcome is correct.
- If outcome is wrong, correct it before it’s official.

Two distinct kinds/stages of audit

1. compliance audit: seek affirmative evidence that the audit trail is sufficiently accurate and intact to reflect the correct outcome.
   - check generation and curation of the trail (ballot accounting, chain of custody, etc.)
   - gives qualitative evidence, like legal evidence
   - if evidence is not convincing, abort
2. materiality audit: seek evidence about whether any errors in recording, transportation, tabulation, reporting that occurred were material, i.e., changed the outcome
   - relies on audit trail: no point if compliance audit fails
   - strategic H2E examination of portions of audit trail ISO convincing evidence that outcome is right
   - gives quantitative, statistical evidence
   - absent convincing evidence, count all votes by hand
   - if audit trail is adequate, that reveals the right outcome

Risk-Limiting Materiality Audits

- Guaranteed minimum chance of correcting the outcome if the outcome is wrong
- Minimum is over all ways the outcome could be wrong: random error, equipment failure, fraud
- Not one method: property of some audits
- Able to tolerate some errors and some deficiencies in the audit trail (fewer unicorns!)

Connection to Statistics

- Formalize audit as sequential statistical hypothesis test
  Null hypothesis: outcome is wrong
  Type I error: conclude outcome is right when it is wrong
  Risk: Chance of Type I error
- Generally test sufficient condition
- Outcome is certainly right if mean of a bounded population is ≤ 1
- Nonparametric test about the mean using some kind of random sample
- The most efficient methods sample individual ballots
- Basic strategies: comparison and ballot-polling

Ballot-polling Audits and Comparison Audits

- Ballot-polling audit:
  Sample ballots until it's clearly pointless to continue:
  looking at the rest would confirm original outcome
  Like an exit poll—but of ballots, not voters
  Soup analogy
- Comparison audit:
  1. Commit to vote subtotals (or CVRs), e.g., precinct-level results
  2. Check that the subtotals add up exactly to contest results
  3. Check subtotals by hand until there is strong evidence the outcome is right
- In general, efficient to let sample size be random:
  audit until evidence is convincing. Size depends on data
- Multiplicity matters: tests, candidates, contests

Tradeoffs

- Ballot polling audit
  - Virtually no set-up costs
  - Requires nothing of voting system
  - Need a ballot manifest to draw sample
  - Preserves voter anonymity except possibly for sampled ballots
  - Requires more counting than ballot-level comparison audit
  - Does not check tabulation: outcome could be right because errors cancel
- Comparison audit
  - Heavy demands on voting system for reporting and data export
  - Requires LEO to commit to subtotals
  - Requires ability to retrieve ballots that correspond to CVRs or subtotals
  - May compromise voter privacy
  - Most efficient (ballot-level) not possible w/ current systems: requires rescan
  - Checks tabulation (but not for transitive audits unless subtotals are cross checked as well)

Pilot Risk-Limiting Audits

- 17 pilot audits in CA, CO, and OH; another 13 planned.
- EAC funding for pilots in CA and CO and Cuyahoga County, OH
- CO has law; CA has pilot law
- simple measures, super-majority, multi-candidate, vote-for-n
- multiple contests audited simultaneously with one sample
- contest sizes: 200 ballots to 121,000 ballots
- counting burden: 16 ballots to 7,000 ballots
- cost per audited ballot: nil to about $0.55
- several jurisdictions have audited on their own—no statistician required

What hasn't been tried?

- Cross-jurisdictional contests (planning for Ohio in 2013)
- IRV/RCV/STV (Victoria? Luxembourg? no-go in San Francisco)

Ballot-polling Audits are often Cheap for Big Contests

255 state-level U.S. presidential contests, 1992–2011, 10% risk limit:
BPA expected to examine fewer than 308 ballots for half the contests.

Work expands as margins shrink, but we could get a lot of election integrity at low cost—with any paper-based system.

Ballot-Polling Audit, 2 Candidates, 10% Risk Limit

Winner's	Ballots drawn		
True Share	median	90th percentile	Mean
70%	22	60	30
65%	38	108	53
60%	84	244	119
55%	332	974	469
53%	914	2,700	1,294
52%	2,051	6,053	2,900
51%	8,157	24,149	11,556
50.5%	32,547	96,411	46,126

Very simple rules and tools for ballot-level audits

Important that calculations be simple and reproducible by observers.
Have approaches easy enough for pencil and paper.

- Comparison: At 10% risk, need 5/margin ballots if no errors are found
  Sample until #good $+\alpha_1 \cdot$#under $-\alpha_2 \cdot$#over $> \alpha_3$
- Ballot-polling: sample until $\alpha_1^{\omega}\alpha_2^{\ell} < \rho$
  $\forall$(winner, loser) pairs.

E2E and paper-based EBE

- Goal of both is to have convincing evidence that outcomes are right—or know that the evidence isn't convincing
- Differ in the nature of evidence, in who generates the evidence, in whom voters need to trust, and for what they must be trusted
- Also differ in ability to recover from corruption of portions of the evidence trail
- Voters, public, and elections officials have different roles in that process in E2E and paper-based EBE
- Examine differences and impact on strength of evidence and anonymity of votes
- Suggest ways to combine and to make E2E more resilient and to parallelize trust requirements

E2E

- Focus on public bulletin-board systems
- Voter can obtain strong evidence that her vote was cast as intended and counted as cast, and that all posted ballots were correctly tabulated
- Enforce vote anonymity using cryptography and procedures (voter cannot prove to anyone how she voted)
- Aggregate votes using homomorphic encryption or mixnet
- Protect voter privacy using randomized threshold public key encryption (requires collusion among officials to break anonymity)

E2E: Typical Assumptions

- "Enough" voters challenge crypto that there's a big chance any problem will be discovered
- "Enough" voters/helpers check PBB that there's a big chance any problem (missing ballots, ballot-stuffing) will be discovered
- If a problem is discovered, it will be reported to the right entity—which will do The Right Thing.
- Voters are not attackers

EBE

- Focus on paper-based systems with compliance & risk-limiting audits
- Voters can obtain strong evidence that vote was cast as intended
- Auditors can obtain strong evidence that outcomes are correct
- Enforce anonymity through equipment and procedures
- Small lapses can break anonymity to elections officials
- Some proposals (e.g., posting digital images of all ballots) could break anonymity to the public

H2E: Typical Assumptions

- Audit trail accurate enough, complete enough, curated well
- LEOs know how many ballots there are and where they are
- If ballots fell off (or on) the truck, LEO will notice and do The Right Thing
- LEOs, pollworkers, auditors trustworthy
- Voter intent discernable from H2E inspection of ballot
- If a ballot has been tampered with, H2E will notice
- Audit gives “adequate” scrutiny to find outcome-changing problems
- If audit finds problems, LEO will do the right thing.

Tradeoffs

	E2E			H2E		
issue	trust	difficulty	how	trust		how
own CAI	self	hard	CHO	self	easy	read
others' CAI	others	hard	CHO	others	easy	read
recorded as cast	self	easy	check BB	LEO/AUD	easy	audit
own CAC	self/public	hard	sum BB	LEO/AUD	easy	audit
others' CAC	self/public	hard	AUD	easy		
authorized voters	self/public/LEO	hard	var.	LEO	easy	regis

CAI: cast as intended

CAC: counted as cast

CHO: cut-and-choose or Benaloh challenge

AUD: auditors

chain of custody versus direct visibility

definition of "any voter"

Outcomes are what matters

Outcome: who won, how many seats each party got, etc.

- Cast as intended, recorded as cast, counted as recorded, reported as counted all sideways
- I don't just care whether my vote counted:
  I want strong evidence that the outcome is right, or an admission that no such evidence is forthcoming
- In the latter case, I want a new election

How can we make an E2E system more resilient and parallelize trust?

- Basic E2E like tamper-evident seal: SI, not SSI
- Can tell that something went wrong—if there's enough scrutiny—but not how badly; generally can't recover
- Tamper-evidence v damage estimate
- Want quantification, not mere detection—& to limit false alarms
- How can we enhance basic strategy to
  - ensure there's enough scrutiny
  - facilitate recovery from errors
  - make it harder to mount a "denial-of-election attack," e.g., from malicious challenges using conterfeit receipts?
  - is there an approach that lets the LEO safely ignore some claims?

Prêt à Voter: Lead Example (Ryan, extended by many)

- To be deployed in Victoria in 11/2014 for rather complex STV
- Original version: pre-printed ballots auditable for correctness
- Because of ballot complexity in Victoria, print-on-demand instead
- Here, examine "traditional" Prêt à Voter
- Requires several unicorns: PBB, people checking PBB hash chains, voters checking crypto, voters checking PBB
- Not best E2E protocol for preventing ballot-box stuffing (not hard to add names to PBB and/or to crypto-votes, but "spooky")

Prêt à Voter: Threats and Vulnerabilities

- Threats to anonymity
- Threats to integrity
- Attack modes:
  - Chain voting
  - Italian attack
  - Randomization attacks
  - Voter keeps both sides of the ballot
  - Ψ-attacks on perceptions of anonymity
  - LEO peeks at printed ballots
  - Bad printing, bad crypto, bad PBB …
  - False claims that PBB is missing cryptovotes: “crying wolf”

Vulnerabilities Prêt à Voter and H2E have in common

- Rely on LEO to determine eligibility
- Rely on LEO to present correct ballot style, in usable format
- Rely on LEO to provide enough ballots, staff, etc.
- Rely on LEO to protect voter anonymity (with Prêt à Voter, mustn't peek at unvoted ballots)
- Rely on LEO to prevent ballot-box stuffing
- What else?

Blending Prêt à Voter with RLA: Sketch

- Distributed construction of ballots with candidates in random order π_i for ballot i
- Commit to two encryptions of each π_i: $\mathrm{PK}_P(\pi_i)$, $\mathrm{PK}_T(\pi_i)$ on PBB, signed by LEO
- Ballot is opscan with candidates in random order π_i.
  Ballot perforated in the middle so candidate list can be separated from bubbles.
  $PKT(\pi_i)$ or an equivalent identifier is printed on the bubble side.
  RHS has carbon
- Printer decrypts $\mathrm{PK}_P(\pi_i)$ and prints ballots with $\mathrm{PK}_T(\pi_i)$, etc. (Signed?)
- LEO/Auditors publicly audit random sample of ballots to confirm that π_i agrees with $\mathrm{PK}_T(\pi_i)$ and check signature.

Prêt à Voter: Sketch, part 2

- Known number of ballots delivered to each polling place.
- Rely on chain of custody to ensure ballots delivered to polling place are authentic and that permutation stays secret.
- Polling-place challenges/audits of encryption and signature; challenged/spoiled ballots kept by pollworkers and returned to LEO, possibly publish decryptions on PBB.
- Voter separates candidate side from bubble side; deposits bubble/$PKT(\pi_i)$ side in ballot box or scanner. Either receives original or copy.
- Candidate side is destroyed or retained by voting system (How to enforce)?

Prêt à Voter: Sketch, part 3

- (mark, $\mathrm{PK}_{\mathcal{T}}(\pi_i)$) pair is used to update the corresponding PBB entry as voted in that way.
  An attempt to update the same PBB entry twice or to update a nonexistent or known-audited entry throws an alarm.
- Poll-closing procedures:
  - Return unused, challenged, & spoiled ballots to LEO
  - Return signed pollbooks to LEO
  - Return electronic data to LEO
  - Return duplicate RHS to LEO for audit.

Prêt à Voter: Sketch, part 4

- Update PBB if that isn't done at polling place; otherwise, check electronic data against PBB.
- Sanity/integrity checks:
  - LEO reconciles pollbook signatures and accounts for ballots: voted, unused, etc.
  - LEO checks that inequalities are satisfied.
  - Mark challenged / spoiled PBB entries and reveal crypto for those.
- Open period for public challenge for receipts missing from PBB.
- Re-encryption mixnet-based tally; NIZKP proof that the mixnet is sound; proof that the decryption is sound.
- Check soundness of PBB hash chain

Prêt à Voter: Sketch, part 5

- Ballot accounting of returned RHS v PBB ballots
- Compliance audit for curation/integrity of RHSs
- Sequential audit to test hypothesis that extra PBB entries + discrepant RHSs cannot account for margin.
  (Reported margin is known at this point.)
  Could include voter-reported missing cryptovotes in the audit.

What does this buy?

- Don’t have to rely on “adequate number” of voters checking receipts “as if at random”
- Independent measurement of rate of missing receipts. Quantitative test for crying wolf.

STAR-Vote

- Travis County, TX. Dan Wallach (lead), Josh Benaloh, Mike Byrne, Bryce Eakin, Phil Kortum, Neal McBurnett, Olivier Pereira, PBS, and Travis Elections staff
- Combine crypto with paper; best of E2E + H2E "belt and suspenders" voting system
- Might lose E2E property for some voters, but keep resilient canvass framework
- Also protects against loss of some paper or loss/corruption of some crypto-data

STAR-Vote w/o crypto details (ask Olivier!)

- Voter interacts w DRE-like device that records but does not cast cryptovotes. Crypto separable by contest.
- Device prints plaintext selections with nonce and 1-d barcode, and crypto-receipt. Benaloh challenge.
- When printed selections go into smart ballot box: box recognizes barcode and flags stored cryptovote as cast
- Voters/helpers can check PBB for cryptovote and can check tally
- LEO and auditors can check that #receipts ≈ #cryptovotes
- System commits to mapping from per-contest votes to ballot ID

STAR-Vote, part 2

- Mapping allows auditors to select per-contest vote and identify corresponding paper (uniqueness of nonce w/i batch verified)
- RLA compares decrypted selection to plaintext on ballot; escalates to full hand count of paper absent strong evidence outcome is right

Conclusions

- It’s about outcomes primarily
- E2E generally “brittle”(at least as specified)
- Can combine E2E and H2E to get more resilience
- Also helps parallelize trust requirements
- Belt and braces!