Exh. 7

Accepted for publication in *Election Law Journal*

Ballot-Marking Devices (BMDs) Cannot Assure the Will of the Voters

Andrew W. Appel[†]
Princeton University

Richard A. DeMillo[†]
Georgia Tech

Philip B. Stark[†]
Univ. of California, Berkeley

February 14, 2020

Abstract

The complexity of U.S. elections usually requires computers to count ballots—but computers can be hacked, so election integrity requires a voting system in which paper ballots can be recounted by hand. However, paper ballots provide no assurance unless they accurately record the votes as expressed by the voters.

Voters can express their intent by indelibly hand-marking ballots, or using computers called ballot-marking device (BMDs). Voters can make mistakes in expressing their intent in either technology, but only BMDs are also subject to hacking, bugs, and misconfiguration of the software that prints the marked ballots. Most voters do not review BMD-printed ballots, and those who do often fail to notice when the printed vote is not what they expressed on the touchscreen. Furthermore, there is no action a voter can take to demonstrate to election officials that a BMD altered their expressed votes, nor is there a corrective action that election officials can take if notified by voters—there is no way to deter, contain, or correct computer hacking in BMDs. These are the essential security flaws of BMDs.

Risk-limiting audits can assure that the votes recorded on paper ballots are tabulated correctly, but no audit can assure that the votes on paper are the ones expressed by the voter on a touchscreen: Elections conducted on current BMDs cannot be confirmed by audits. We identify two properties of voting systems, *contestability* and *defensibility*, necessary for audits to confirm election outcomes. No available EAC-certified BMD is contestable or defensible.

[†]Authors are listed alphabetically; they contributed equally to this work.

1 Introduction: Criteria for Voting Systems

Elections for public office and on public questions in the United States or any democracy must produce outcomes based on the votes that voters *express* when they indicate their choices on a paper ballot or on a machine. Computers have become indispensable to conducting elections, but computers are vulnerable. They can be hacked—compromised by insiders or external adversaries who can replace their software with fraudulent software that deliberately miscounts votes—and they can contain design errors and bugs—hardware or software flaws or configuration errors that result in mis-recording or mis-tabulating votes. Hence there must be some way, *independent* of any software in any computers, to ensure that reported election outcomes are correct, i.e., consistent with the expressed votes as intended by the voters.

Voting systems should be *software independent*, meaning that "an undetected change or error in its software cannot cause an undetectable change or error in an election outcome" [30, 31, 32]. Software independence is similar to tamper-evident packaging: if somebody opens the container and disturbs the contents, it will leave a trace.

The use of software-independent voting systems is supposed to ensure that if someone fraudulently hacks the voting machines to steal votes, we'll know about it. But we also want to know *the true outcome* in order to avoid a do-over election.[1] A voting system is *strongly software independent* if it is software independent and, moreover, a detected change or error in an election outcome (due to change or error in the software) can be corrected using only the ballots and ballot records of the current election [30, 31]. Strong software independence combines tamper evidence with a kind of resilience: there's a way to tell whether faulty software caused a problem, and a way to recover from the problem if it did.

Software independence and *strong software independence* are now standard terms in the analysis of voting systems, and it is widely accepted that voting systems should be software independent. Indeed, version 2.0 of the Voluntary Voting System Guidelines (VVSG 2.0) incorporates this principle [11].

But as we will show, these standard definitions are incomplete and inadequate, because the word *undetectable* hides several important questions: *Who* detects the change or error in an election outcome? How can a person *prove* that she has detected an er-

[1] Do-overs are expensive; they may delay the inauguration of an elected official; there is no assurance that the same voters will vote in the do-over election as voted in the original; they decrease public trust. And if the do-over election is conducted with the same voting system that can only detect but not correct errors, then there may need to be a do-over of the do-over, *ad infinitum.*

ror? *What happens* when someone detects an error—does the election outcome remain erroneous? Or conversely: How can an election administrator *prove* that the election outcome not been altered, or prove that the correct outcome was recovered if a software malfunction was detected? The standard definition does not distinguish evidence available to an election official, to the public, or just to a single voter; nor does it consider the possibility of false alarms.

Those questions are not merely academic, as we show with an analysis of ballot-marking devices. Even if some *voters* "detect" that the printed output is not what they expressed to the BMD—even if some of *those* voters report their detection to election officials—there is no mechanism by which the *election official* can "detect" whether a BMD has been hacked to alter election outcomes. The questions of *who detects, and then what happens,* are critical—but unanswered by the standard definitions.

We will define the terms *contestable* and *defensible* to better characterize properties of voting systems that make them acceptable for use in public elections.[2]

A voting system is *contestable* if an undetected change or error in its software that causes a change or error in an election outcome can always produce *public* evidence that the outcome is untrustworthy. For instance, if a voter selected candidate A on the touchscreen of a BMD, but the BMD prints candidate B on the paper ballot, then this A-vs-B evidence is available to the individual voter, but the voter cannot demonstrate this evidence to anyone else, since nobody else saw—nor should have seen—where the voter touched the screen.[3] Thus, the voting system does not provide a way for the voter who observed the misbehavior to prove to anyone else that there was a problem, even if the problems altered the reported outcome. Such a system is therefore not *contestable*.

While the definition of software independence might allow evidence available only to individual voters as "detection," such evidence does not suffice for a system to be contestable. Contestibility is software independence, plus the requirement that "detect" implies "can generate public evidence." "Trust me" does not count as public evidence. If a voting system is not contestable, then problems voters "detect" might never see the light of day, much less be addressed or corrected.[4]

[2]There are other notions connected to contestability and defensibility, although essentially different: Benaloh et al. [6] define a *P-resilient canvass framework*, *personally verifiable P-resilient canvass framework*, and *privacy-perserving personally verifiable P-resilient canvass frameworks*.

[3]See footnote 17.

[4]If voters are the only means of detecting and quantifying the effect of those problems—as they are for BMDs—then in practice the system is not strongly software independent. The reason is that, as we will show, such claims by (some) voters *cannot* correct software-dependent changes to other voters' ballots, and *cannot* be used as the basis to invalidate or correct an election outcome. Thus, BMD-based

Similarly, while strong software independence demands that a system be able to report the correct outcome even if there was an error or alteration of the software, it does not require *public evidence* that the (reconstructed) reported outcome is correct. We believe, therefore, that voting systems must also be *defensible*. We say that a voting system is defensible if, when the reported electoral outcome is correct, it is possible to generate convincing public evidence that the reported electoral outcome is correct—despite any malfunctions, software errors, or software alterations that might have occurred. If a voting system is not defensible, then it is vulnerable to "crying wolf": malicious actors could claim that the system malfunctioned when in fact it did not, and election officials will have no way to prove otherwise.

By analogy with *strong software independence*, we define: A voting system is *strongly defensible* if it is defensible and, moreover, a detected change or error in an election outcome (due to change or error in the software) can be corrected (with convincing public evidence) using only the ballots and ballot records of the current election.

In short, a system is contestable if it can generate public evidence of a problem whenever a reported outcome is wrong, while a system is defensible if it can generate public evidence whenever a reported outcome is correct—despite any problems that might have occurred. Contestable systems are publicly tamper-evident; defensible systems are publicly, demonstrably resilient.

Defensibility is a key requirement for *evidence-based elections* [39]: defensibility makes it possible in principle for election officials to generate convincing evidence that the reported winners really won—if the reported winners did really win. (We say an election *system* may be defensible, and an *election* may be evidence-based; there's much more *process* to an election than just the choice of system.)

Examples. The only known practical technology for contestable, strongly defensible voting is a system of *hand-marked paper ballots*, kept demonstrably physically secure, counted by machine, audited manually, and recountable by hand.[5] In a hand-marked paper ballot election, ballot-marking software cannot be the source of an error or change-of-election-outcome, because no software is used in marking ballots. Ballot-scanning-and-counting software can be the source of errors, but such errors can be

election systems are not even (weakly) software independent, unless one takes "detection" to mean "somebody claimed there was a problem, with no evidence to support that claim."

[5]The election must also generate convincing evidence that physical security of the ballots was not compromised, and the audit must generate convincing public evidence that the audit itself was conducted correctly.

detected and corrected by audits.

That system is *contestable:* if an optical scan voting machine reports the wrong outcome because it miscounted (because it was hacked, misprogrammed, or miscalibrated), the evidence is *public*: the paper ballots, recounted before witnesses, will not match the claimed results, also witnessed. It is *strongly defensible:* a recount before witnesses can demonstrate that the reported outcome is correct, or can find the correct outcome if it was wrong—and provide public evidence that the (reconstructed) outcome is correct. See Section 4 for a detailed analysis.

Over 40 states now use some form of paper ballot for most voters [19]. Most of the remaining states are taking steps to adopt paper ballots. But *not all voting systems that use paper ballots are equally secure*.

Some are not even software independent. Some are software independent, but not strongly software independent, contestable, or defensible. In this report we explain:

- *Hand-marked paper ballot* systems are the only practical technology for contestable, strongly defensible voting systems.
- *Some ballot-marking devices (BMDs)* can be software independent, but they not strongly software independent, contestable, or defensible. Hacked or misprogrammed BMDs can alter election outcomes undetectably, so elections conducted using BMDs cannot provide public evidence that reported outcomes are correct. If BMD malfunctions are detected, there is no way to determine who really won. Therefore BMDs should not be used by voters who are able to mark an optical-scan ballot with a pen.
- *All-in-one BMD or DRE+VVPAT voting machines* are not software independent, contestable, or defensible. They should not be used in public elections.

2 Background

We briefly review the kinds of election equipment in use, their vulnerability to computer hacking (or programming error), and in what circumstances risk-limiting audits can mitigate that vulnerability.

Voting equipment

Although a voter may form an intention to vote for a candidate or issue days, minutes, or seconds before actually casting a ballot, that intention is a psychological state that cannot be directly observed by anyone else. Others can have access to that intention through what the voter (privately) *expresses* to the voting technology by interacting with it, e.g., by making selections on a BMD or marking a ballot by hand.[6] Voting systems must accurately record the vote as the voter *expressed* it.

With a *hand-marked paper ballot optical-scan* system, the voter is given a paper ballot on which all choices (candidates) in each contest are listed; next to each candidate is a *target* (typically an oval or other shape) which the voter marks with a pen to indicate a vote. Ballots may be either preprinted or printed (unvoted) at the polling place using *ballot on demand* printers. In either case, the voter creates a tamper-evident record of intent by marking the printed paper ballot with a pen.

Such hand-marked paper ballots may be scanned and tabulated at the polling place using a *precinct-count optical scanner* (PCOS), or may be brought to a central place to be scanned and tabulated by a *central-count optical scanner* (CCOS). Mail-in ballots are typically counted by CCOS machines.

After scanning a ballot, a PCOS machine deposits the ballot in a secure, sealed ballot box for later use in recounts or audits; this is *ballot retention*. Ballots counted by CCOS are also retained for recounts or audits.[7]

Paper ballots can also be hand counted, but in most jurisdictions (especially where there are many contests on the ballot) this is hard to do quickly; Americans expect election-night reporting of unofficial totals. Hand counting—i.e., manually determining votes directly from the paper ballots—is appropriate for audits and recounts.

A *ballot-marking device* (BMD) provides a computerized user interface that presents

[6] We recognize that voters make mistakes in expressing their intentions. For example, they may misunderstand the layout of a ballot or express an unintended choice through a perceptual error, inattention, or lapse of memory. The use of touchscreen technology does not necessarily correct for such user errors, as every smartphone user who has mistyped an important text message knows. Poorly designed ballots, poorly designed touchscreen interfaces, and poorly designed assistive interfaces increase the rate of error in voters' expressions of their votes. For the purposes of this report, we assume that properly engineered systems seek to minimize such usability errors.

[7] Regulations and procedures governing custody and physical security of ballots are uneven and in many cases inadequate, but straightforward to correct because of decades of development of best practices.

the ballot to voters and captures their expressed selections—for instance, a touchscreen interface or an assistive interface that enables voters with disabilities to vote independently. Voter inputs (expressed votes) are recorded electronically. When a voter indicates that the ballot is complete and ready to be cast, the BMD prints a paper version of the electronically marked ballot. We use the term *BMD* for devices that mark ballots but do not tabulate or retain them, and *all-in-one* for devices that combine ballot marking, tabulation, and retention into the same paper path.

The paper ballot printed by a BMD may be in the same format as an optical-scan form (e.g., with ovals filled as if by hand) or it may list just the names of the candidate(s) selected in each contest. The BMD may also encode these selections into barcodes or QR codes for optical scanning. We discuss issues with barcodes later in this report.

An *all-in-one touchscreen voting machine* combines computerized ballot marking, tabulation, and retention in the same paper path. All-in-one machines come in several configurations:

- DRE+VVPAT machines—direct-recording electronic (DRE) voting machines with a voter-verifiable paper audit trail (VVPAT)—provide the voter a touchscreen (or other) interface, then print a paper ballot that is displayed to the voter under glass. The voter is expected to review this ballot and approve it, after which the machine deposits it into a ballot box. DRE+VVPAT machines do not contain optical scanners; that is, they do not read what is marked on the paper ballot; instead, they tabulate the vote directly from inputs to the touchscreen or other interface.
- BMD+Scanner all-in-one machines[8] provide the voter a touchscreen (or other) interface to input ballot choices and print a paper ballot that is ejected from a slot for the voter to inspect. The voter then reinserts the ballot into the slot, after which the all-in-one BMD+scanner scans it and deposits it into a ballot box. Or, some BMD+Scanner all-in-one machines display the paper ballot behind plexiglass for the voter to inspect, before mechanically depositing it into a ballot box.

Opscan+BMD with separate paper paths. At least one model of voting machine (the Dominion ICP320) contains an optical scanner (opscan) and a BMD in the same cabinet,[9] so that the optical scanner and BMD-printer are not in the same paper path; no possible configuration of the software could cause a BMD-marked ballot to be deposited in the ballot box without human handling of the ballot. We do not classify this as an *all-in-one* machine.

[8] Some voting machines, such as the ES&S ExpressVote, can be configured as either a BMD or a BMD+Scanner all-in-one. Others, such as the ExpressVoteXL, work only as all-in-one machines.

[9] More precisely, the ICP320 optical scanner and the BMD audio+buttons interface are in the same cabinet, but the printer is a separate box.

Hacking

There are many forms of computer hacking. In this analysis of voting machines we focus on the alteration of voting machine software so that it miscounts votes or mismarks ballots to alter election outcomes. There are many ways to alter the software of a voting machine: a person with physical access to the computer can open it and directly access the memory; one can plug in a special USB thumbdrive that exploits bugs and vulnerabilities in the computer's USB drivers; one can connect to its WiFi port or Bluetooth port or telephone modem (if any) and exploit bugs in those drivers, or in the operating system.

"Air-gapping" a system (i.e., never connecting it to the Internet nor to any other network) does not automatically protect it. Before each election, election administrators must transfer a *ballot definition* into the voting machine by inserting a *ballot definition cartridge* that was programmed on election-administration computers that may have been connected previously to various networks; it has been demonstrated that vote-changing viruses can propagate via these ballot-definition cartridges [18].

Hackers might be corrupt insiders with access to a voting-machine warehouse; corrupt insiders with access to a county's election-administration computers; outsiders who can gain remote access to election-administration computers; outsiders who can gain remote access to voting-machine manufacturers' computers (and "hack" the firmware installed in new machines, or the firmware updates supplied for existing machines), and so on. Supply-chain hacks are also possible: the hardware installed by a voting system vendor may have malware pre-installed by the vendor's component suppliers.[10]

Computer systems (including voting machines) have so many layers of software that it is impossible to make them perfectly secure [24, pp. 89–91]. When manufacturers of voting machines use the best known security practices, adversaries may find it more difficult to hack a BMD or optical scanner—but not impossible. Every computer in every critical system is vulnerable to compromise through hacking, insider attacks or exploiting design flaws.

[10]Given that many chips and other components are manufactured in China and elsewhere, this is a serious concern. Carsten Schürmann has found Chinese pop songs on the internal memory of voting machines (C. Schürmann, personal communication, 2018). Presumably those files were left there accidentally—but this shows that malicious code *could* have been pre-installed deliberately, and that neither the vendor's nor the election official's security and quality control measures discovered and removed the extraneous files.

Election assurance through risk-limiting audits

To ensure that the reported electoral outcome of each contest corresponds to what the voters expressed, the most practical known technology is a *risk-limiting audit* (RLA) of trustworthy paper ballots [35, 36, 23]. The National Academies of Science, Engineering, and Medicine, recommend routine RLAs after every election [24], as do many other organizations and entities concerned with election integrity.[11]

The *risk limit* of a risk-limiting audit is the maximum chance that the audit will not correct the reported electoral outcome, if the reported outcome is wrong. "Electoral outcome" means the political result—who or what won—not the exact tally. "Wrong" means that the outcome does not correspond to what the voters expressed.

A RLA involves manually inspecting randomly selected paper ballots following a rigorous protocol. The audit stops if and when the sample provides convincing evidence that the reported outcome is correct; otherwise, the audit continues until every ballot has been inspected manually, which reveals the correct electoral outcome if the paper trail is trustworthy. RLAs protect against vote-tabulation errors, whether those errors are caused by failures to follow procedures, misconfiguration, miscalibration, faulty engineering, bugs, or malicious hacking.[12]

The risk limit should be determined as a matter of policy or law. For instance, a 5% risk limit means that, if a reported outcome is wrong solely because of tabulation errors, there is at least a 95% chance that the audit procedure will correct it. Smaller risk limits give higher confidence in election outcomes, but require inspecting more ballots, other things being equal. RLAs never revise a correct outcome.

RLAs can be very efficient, depending in part on how the voting system is designed and how jurisdictions organize their ballots. If the computer results are accurate, an efficient RLA with a risk limit of 5% requires examining just a few—about 7 divided by the margin—ballots selected randomly from the contest.[13] For instance, if the margin of victory is 10% and the results are correct, the RLA would need to examine about $7/10\% = 70$ ballots to confirm the outcome at 5% risk. For a 1% margin, the RLA would need to examine about $7/1\% = 700$ ballots. The sample size does not depend

[11] Among them are the Presidential Commission on Election Administration, the American Statistical Association, the League of Women Voters, and Verified Voting Foundation.

[12] RLAs do not protect against problems that cause BMDs to print something other than what was shown to the voter on the screen, nor do they protect against problems with ballot custody.

[13] Technically, it is the *diluted margin* that enters the calculation. The diluted margin is the number of votes that separate the winner with the fewest votes from the loser with the most votes, divided by the number of ballots cast, including undervotes and invalid votes.

much on the total number of ballots cast in the contest, only on the margin of the winning candidate's victory.

RLAs assume that a full hand tally of the paper trail would reveal the correct electoral outcomes: the paper trail must be trustworthy. Other kinds of audits, such as *compliance audits* [6, 23, 39, 37] are required to establish whether the paper trail itself is trustworthy. Applying an RLA procedure to an untrustworthy paper trail cannot limit the risk that a wrong reported outcome goes uncorrected.

Properly preserved hand-marked paper ballots ensure that expressed votes are identical to recorded votes. But BMDs might not record expressed votes accurately, for instance, if BMD software has bugs, was misconfigured, or was hacked: BMD printout is not a trustworthy record of the expressed votes. Neither a compliance audit nor a RLA can possibly check whether errors in recording expressed votes altered election outcomes. RLAs that rely on BMD output therefore cannot limit the risk that an incorrect reported election outcome will go uncorrected.

A paper-based voting system (such as one that uses optical scanners) is systematically more secure than a paperless system (such as DREs) *only if the paper trail is trustworthy and the results are checked against the paper trail using a rigorous method such as an RLA or full manual tally*. If it is possible that error, hacking, bugs, or miscalibration caused the recorded-on-paper votes to differ from the expressed votes, an RLA or even a full hand recount cannot not provide convincing public evidence that election outcomes are correct: such a system cannot be *defensible*. In short, paper ballots provide little assurance against hacking if they are never examined or if the paper might not accurately reflect the votes expressed by the voters.

3 (Non)Contestability/Defensibility of BMDs

A BMD-generated paper trail is not a reliable record of the vote expressed by the voter. Like any computer, a BMD (or a DRE+VVPAT) is vulnerable to bugs, misconfiguration, hacking, installation of unauthorized (fraudulent) software, and alteration of installed software.

If a hacker sought to steal an election by altering BMD software, what would the hacker program the BMD to do? In cybersecurity practice, we call this the *threat model*.

The simplest threat model is this one: In some contests, not necessarily top-of-the-ticket, change a small percentage of the votes (such as 5%).

In recent national elections, analysts have considered a candidate who received 60% of the vote to have won by a landslide. Many contests are decided by less than a 10% margin. Changing 5% of the votes can change the margin by 10%, because "flipping" a vote for one candidate into a vote for a different candidate changes the difference in their tallies—i.e., the margin—by 2 votes. If hacking or bugs or misconfiguration could change 5% of the votes, that would be a very significant threat.

Although public and media interest often focus on top-of-the-ticket races such as President and Governor, elections for lower offices such as state representatives, who control legislative agendas and redistricting, and county officials, who manage elections and assess taxes, are just as important in our democracy. Altering the outcome of smaller contests requires altering fewer votes, so fewer voters are in a position to notice that their ballots were misprinted. And most voters are not as familiar with the names of the candidates for those offices, so they might be unlikely to notice if their ballots were misprinted, even if they checked.

Research in a real polling place in Tennessee during the 2018 election, found that half the voters *didn't look at all* at the paper ballot printed by a BMD, even when they were holding it in their hand and directed to do so while carrying it from the BMD to the optical scanner [14]. Those voters who did look at the BMD-printed ballot spent *an average of 4 seconds* examining it to verify that the eighteen or more choices they made were correctly recorded. That amounts to 222 milliseconds per contest, barely enough time for the human eye to move and refocus under perfect conditions and not nearly enough time for perception, comprehension, and recall [28]. A study by other researchers [8], in a simulated polling place using real BMDs deliberately hacked to alter one vote on each paper ballot, found that only 6.6% of voters told a pollworker something was wrong.[14][15] The same study found that among voters who examined their hand-marked ballots, half were unable to recall key features of ballots cast moments before, a prerequisite step for being able to recall their own ballot choices. This finding is broadly consistent with studies of effects like "change blindness" or "choice blindness," in which human subjects fail to notice changes made to choices

[14] You might think, "the voter really *should* carefully review their BMD-printed ballot." But because the scientific evidence shows that voters *do not* [14] and cognitively *cannot* [17] perform this task well, legislators and election administrators should provide a voting system that counts the votes *as voters express them*.

[15] Studies of voter confidence about their ability to verify their ballots are not relevant: in typical situations, subjective confidence and objective accuracy are at best weakly correlated. The relationship between confidence and accuracy has been studied in contexts ranging from eyewitness accuracy [9, 13, 42] to confidence in psychological clinical assessments [15] and social predictions [16]. The disconnect is particularly severe at high confidence. Indeed, this is known as "the overconfidence effect." For a lay discussion, see *Thinking, Fast and Slow* by Nobel economist Daniel Kahnemann [21].

made only seconds before [20].

Suppose, then, that 10% of voters examine their paper ballots carefully enough to even *see* the candidate's name recorded as their vote for legislator or county commissioner. Of those, perhaps only half will remember the name of the candidate they intended to vote for.[16]

Of those who notice that the vote printed is not the candidate they intended to vote for, what will they think, and what will they do? Will they think, "Oh, I must have made a mistake on the touchscreen," or will they think, "Hey, the machine is cheating or malfunctioning!" There's no way for the voter to know for sure—voters do make mistakes—and there's *absolutely* no way for the voter to prove to a pollworker or election official that a BMD printed something other than what the voter entered on the screen.[17][18]

Either way, polling-place procedures generally advise voters to ask a pollworker for a new ballot if theirs does not show what they intended. Pollworkers should void that BMD-printed ballot, and the voter should get another chance to mark a ballot. Anecdotal evidence suggests that many voters are too timid to ask, or don't know that they have the right to ask, or are not sure whom to ask. Even if a voter asks for a new ballot, training for pollworkers is uneven, and we are aware of no formal procedure for resolving disputes if a request for a new ballot is refused. Moreover, there is no sensible protocol for ensuring that BMDs that misbehave are investigated—nor can there be, as we argue below.

Let's summarize. If a machine alters votes on 5% of the ballots (enabling it to change the margin by 10%), and 10% of voters check their ballots carefully and 50% of the voters who check notice the error, then optimistically we might expect $5\% \times 10\% \times 50\%$ or 0.25% of the voters to request a new ballot and correct their vote.[19] This

[16] We ask the reader, "do you know the name of the most recent losing candidate for county commissioner?" We recognize that some readers of this document *are* county commissioners, so we ask those readers to imagine the frame of mind of their constituents.

[17] You might think, "the voter can prove it by showing someone that the vote on the paper doesn't match the vote onscreen." But that won't work. On a typical BMD, by the time a paper record is printed and ejected for the voter to hold and examine, the touchscreen no longer shows the voter's choice. You might think, "BMDs should be designed so that the choices still show on the screen for the voter to compare with the paper." But a hacked BMD could easily alter the on-screen choices to match the paper, *after* the voter hits the "print" button.

[18] Voters should *certainly not* videorecord themselves voting! That would defeat the privacy of the secret ballot and is illegal in most jurisdictions.

[19] This calculation assumes that the 10% of voters who check are in effect a random sample of voters: voters' propensity to check BMD printout is not associated with their political preferences.

means that the machine will change the margin by 9.75% and get away with it.

In this scenario, 0.25% of the voters, one in every 400 voters, has requested a new ballot. You might think, "that's a form of *detection* of the hacking." But is isn't, as a practical matter: a few individual voters may have detected that there was a problem, but there's no procedure by which this translates into any action that election administrators can take to correct the outcome of the election. Polling-place procedures *cannot correct or deter hacking, or even reliably detect it*, as we discuss next. This is essentially the distinction between a system that is merely software independent and one that is contestable: a change to the software that alters the outcome might generate evidence for an alert, conscientious, individual voter, but it does not generate public evidence that an election official can rely on to conclude there is a problem.

Even if some voters notice that BMDs are altering votes, there's no way to correct the election outcome. That is, BMD voting systems are *not contestable*, *not defensible* (and therefore *not strongly defensible*), and *not strongly software independent*. Suppose a state election official wanted to detect whether the BMDs are cheating, and correct election results, based on actions by those few alert voters who notice the error. What procedures could possibly work against the manipulation we are considering?

1. How about, "If at least 1 in 400 voters claims that the machine misrepresented their vote, void the entire election."[20] No responsible authority would implement such a procedure. A few dishonest voters could collaborate to invalidate entire elections simply by falsely claiming that BMDs changed their votes.
2. How about, "If at least 1 in 400 voters claims that the machine misrepresented their vote, then investigate." Investigations are fine, but then what? The only way an investigation can ensure that the outcome accurately reflects what voters expressed to the BMDs is to void an election in which the BMDs have altered votes and conduct a new election. But how do you know whether the BMDs have altered votes, except based the claims of the voters?[21] Furthermore, the investigation itself would suffer from the same problem as above: how can one

[20]Note that in many jurisdictions, far fewer than 400 voters use a given machine on election day: BMDs are typically expected to serve fewer than 300 voters per day. (The vendor ES&S recommended 27,000 BMDs to serve Georgia's 7 million voters, amounting to 260 voters per BMD [34].) Recall also that the rate 1 in 400 is tied to the amount of manipulation. What if the malware flipped only one vote in 50, instead of 1 vote in 20? That could still change the margin by 4%, but—in this hypothetical—would be noticed by only one voter in 1,000, rather than one in 400. The smaller the margin, the less manipulation it would have taken to alter the electoral outcome.

[21]Forensic examination of the BMD might show that it *was* hacked or misconfigured, but it cannot prove that the BMD *was not* hacked or misconfigured.

distinguish between voters who detected BMD hacking or bugs from voters who just want to interfere with an election?

This is the essential security flaw of BMDs: few voters will notice and promptly report discrepancies between what they saw on the screen and what is on the BMD printout, and even when they do notice, there's nothing appropriate that can be done. Even if election officials are convinced that BMDs malfunctioned, *there is no way to determine who really won*.

Therefore, BMDs should not be used by most voters.

Why can't we rely on pre-election and post-election logic and accuracy testing, or parallel testing? Most, if not all, jurisdictions perform some kind of *logic and accuracy testing* (LAT) of voting equipment before elections. LAT generally involves voting on the equipment using various combinations of selections, then checking whether the equipment tabulated the votes correctly. As the Volkswagen/Audi "Dieselgate" scandal shows, devices can be programmed to behave properly when they are tested but misbehave in use [12]. Therefore, LAT can never prove that voting machines performed properly in practice.

Parallel or "live" testing involves pollworkers or election officials using some BMDs at random times on election day to mark (but not cast) ballots with test patterns, then check whether the marks match the patterns. The idea is that the testing is not subject to the "Dieselgate" problem, because the machines cannot "know" they are being tested on election day. As a practical matter, the number of tests required to provide a reasonable chance of detecting outcome-changing errors is prohibitive, and even then the system is not *defensible.* See Section 6.

Suppose, counterfactually, that it was practical to perform enough parallel testing to guarantee a large chance of detecting a problem if BMD hacking or malfunction altered electoral outcomes. Suppose, counterfactually, that election officials were required to conduct that amount of parallel testing during every election, and that the required equipment, staffing, infrastructure, and other resources were provided. Even then, the system would not be *strongly defensible*; that is, if testing detected a problem, there would be no way to to determine who really won. The only remedy would be a new election.

Don't voters need to check hand-marked ballots, too? It is always a good idea to check one's work, but there is a substantial body of research (e.g., [29]) suggesting

that preventing error as a ballot is being marked is a fundamentally different cognitive task than detecting an error on a previously marked ballot. In cognitively similar tasks, such as proof reading for non-spelling errors, ten percent rates of error detection are common [29, pp 167ff], whereas by carefully attending to the task of correctly marking their ballots, voters apparently can largely avoid marking errors.

A fundamental difference between hand-marked paper ballots and ballot-marking devices is that, with hand-marked paper ballots, voters are responsible for catching and correcting *their own errors*, while if BMDs are used, voters are also responsible for catching *machine errors, bugs, and hacking*. Voters are the *only* people who can detect such problems with BMDs—but, as explained above, if voters do find problems, there's no way they can prove to poll workers or election officials that there were problems and no way to ensure that election officials take appropriate remedial action.

4 Contestability/defensibility of hand-marked opscan

The most widely used voting system in the United States optical-scan counting of hand-marked paper ballots.[22] Computers and computer software are used in several stages of the voting process, and if that software is hacked (or erroneous), then the computers will deliberately (or accidentally) report incorrect outcomes.

- Computers are used to prepare the PDF files from which (unvoted) optical-scan ballots are printed, with ovals (or other targets to be marked) next to the names of candidates. Because the optical scanners respond to the *position on the page,* not the name of the candidate nearest the target, computer software could cheat by reordering the candidates on the page.
- The optical-scan voting machine, which scans the ballots and interprets the marks, is driven by computer software. Fraudulent (hacked) software can deliberately record (some fraction of) votes for Candidate A and votes for Candidate B.
- After the voting machine reports the in-the-precinct vote totals (or, in the case of central-count optical scan, the individual-batch vote totals), computers are used to aggregate the various precincts or batches together. Hacked software could cheat in this addition process.

Protection against any or all of these attacks relies on a system of risk-limiting

[22]The Verifier – Polling Place Equipment – November 2020, https://www.verifiedvoting.org/verifier/, Verified Voting Foundation, fetched February 8, 2020.

audits, along with compliance audits to check that the chain of custody of ballots and paper records is trustworthy. Without such audits, optical-scan ballots (whether hand marked or machine marked) are neither contestable nor defensible.

We analyze the contestability/defensibility of hand-marked optical-scan ballots with respect to each of these threats, assuming a system of RLAs and compliance audits.

- Hacked generation PDFs leading to fraudulently placed ovals. In this case, a change or error in the computer software *can* change the election outcome: on thousands of ballots, voters place a mark next to the name of candidate A, but (because the candidate name has been fraudulently misplaced on the paper), the (unhacked) optical scanner records this as a vote for candidate B. But an RLA will correct the outcome: a human, inspecting and interpreting this paper ballot, will interpret the mark as a vote for candidate A, as the voter intended. The RLA will, with high probability, conclude that the computer-reported election outcome cannot be confirmed, and a full recount must occur. Thus the system is *contestable:* the RLA produces public evidence that the (computer-reported) outcome is untrustworthy. This full recount (in the presence of witnesses, in view of the public) can provide convincing public evidence of its own correctness; that is, the system is *defensible*.
- Hacked optical-scan vote counter, reporting fraudulent vote totals. In this case, a change or error in the computer software *can* change the election outcome: on thousands of ballots, voters place a mark next to the name of candidate A, but the (hacked) optical scanner records this as a vote for candidate B. But an RLA can detect the incorrect outcome (just as in the case above); the system is *contestable*. And a full recount will produce a correct outcome with public evidence: the system is *defensible*.
- Hacked election-management system (EMS), fraudulently aggregating batches. A risk-limiting audit can detect this problem, and a recount will correct it: the system is contestable and defensible. But actually, contestability and defensibility against this attack is even easier and simpler than RLAs and recounts. Most voting machines (including precinct-count optical scanners) print a "results tape" in the polling place, at the close of the polls (in addition to writing their results electronically to a removable memory card). This results tape is (typically) signed by pollworkers and by credentialed challengers, and open to inspection by members of the public, before it is transported (with chain-of custody protections) along with the ballot boxes to a secure central location. The County Clerk or Registrar of Voters can (and in many counties, does) inspect these paper records to verify that they correspond to the precinct-by-precinct machine-reported aggregation. Errors (or fraud) in aggregation can be detected and cor-

rected without the need to inspect individual ballots: the system is contestable and defensible against this class of errors.

5 End-to-end verifiable (E2E-V) systems

In all BMD systems currently on the market, and in all BMD systems certified by the EAC, the printed ballot or ballot summary is the only channel by which voters can verify the correct recording of their ballots, independently of the computers. The analysis in this paper applies to all of those BMD systems.

There is a class of voting systems called "end-to-end verifiable" (E2E-V), which provide an alternate mechanism for voters to verify their votes [7] [2]. The basic idea of an E2E-V system is that a cryptographic protocol encodes the vote; mathematical properties of the cryptographic system allow the voters to verify (probabilistically) that their vote has been accurately counted, but does not compromise secret ballot by allowing voters to prove how they voted. E2E-V systems have not been adopted in public elections (except that Scantegrity was used for municipal elections in Takoma Park, MD in 2009 and 2011).

Each E2E-V system requires its own analysis of contestability/defensibility.

Scantegrity [10] is a system of preprinted optical-scan ballots, counted by conventional precinct-count optical scanners, but with an additional security feature: when the voter fills in an oval with a special pen, the oval is mostly darkened (so it's counted conventionally by the optical scanner), but two-letter code is also revealed that the voter can (optionally) use in the cryptographic protocol. Scantegrity is contestable/defensible, but not because of its E2E-V properties: since it's an add-on to a conventional optical-scan system with hand-marked paper ballots, RLAs and compliance audits can render this system contestable/defensible.

Prêt-à-Voter [33] is the system in which the voter separates the candidate-list from the oval-target list after marking the ballot and before deposit into the optical scanner. This system can be made contestable, with difficulty: the auditing procedure requires participation of the voters in an unintuitive cryptographic challenge. It is not clear that the system is defensible: if this cryptographic challenge proves that the blank ballots

have been tampered with, then no recount can reliably reconstruct the true result with public evidence.

STAR-Vote [5] is a DRE+VVPAT system with a smart ballot box. Voters interact with a device that captures their votes electronically and prints a paper record that voters can inspect, but the electronic votes are held "in limbo" until the paper ballot is deposited in the smart ballot box. The ballot box does not read the votes from the ballot; rather, depositing the ballot tells the system that it has permission to cast the votes it had already recorded from the touchscreen. The claimed advantage of STAR-Vote (and other systems that use the "Benaloh challenge") is that RLAs and ballot-box chain-of-custody are not required in order to obtain software independence. To assure that the E2E-V cryptographic protocol has correctly recorded each vote, the voter can "challenge" the system to prove that the cryptographic encoding of the ballot records the vote actually printed on the paper ballot. To do so, the voter must discard (void) this ballot and vote a fresh ballot; this is because the challenge process reveals the vote to the public, and a voting system must preserve the secrecy of the (cast) ballots. Thus, the voter cannot ensure the correct encoding of their true ballot, but (since STAR-Vote must print the ballot before knowing whether the voter will challenge), the voter can ensure it with any desired *error probability*.

STAR-Vote is software independent but it is not contestable or defensible. The reason is that, while the challenge can produce public evidence that a machine did not accurately encrypt the plaintext vote on the ballot, if the machine prints the wrong plaintext vote and a correct encryption of that incorrect vote, there is no evidence the voter can use to prove that to anyone else.

No E2E-V system is currently certified by the EAC, nor to our knowledge is any such system under review for certification, nor are any of the 5 major voting-machine vendors offering such a system for sale.[23]

[23] Some vendors, notably Scytl, have sold systems advertised as E2E-V in other countries. Those systems were not in fact E2E-V. Moreover, serious security flaws have been found in their implementations. See, e.g., [22].

6 Parallel testing of BMDs

Wallach [41] has proposed (in response to earlier drafts of this paper) that contestability/defensibility failure of BMDs could be mitigated by *parallel testing*, which he also calls "live auditing." Stark [38] has analyzed Wallach's proposal in detail. Here we provide a summary of the proposal and the analyses.

One might like to test each BMD before the election to make sure it's not hacked. Unfortunately, since the computer in a voting machine (including BMDs) has a real-time clock, the software (including fraudulent vote-stealing software) knows whether it's election day or not. Fraudulent software can make sure not to cheat except on election day.

The idea of parallel testing is to have trained auditors test the BMDs, at random times during an actual election: use the BMD to prepare a ballot, inspect that ballot to ensure it's marked correctly, then discard the ballot. The same BMDs in use during the polling will be selected, from time to time, for such test, right there in the polling places.

If the BMDs cheat with uniform random probability p, and if the BMD cannot distinguish an auditor from an ordinary voter, then after n random audits the probability of detecting the malware is $1 - (1 - p)^n$. If $p = 5\%$ and $n = 240$, then the probability of detection is 91%.

Unfortunately, the attacker is not constrained to cheat with uniform random probability; or, to put it another way, BMD malware may indeed be able to distinguish auditors from ordinary voters. Stark [38] discusses many ways in which the "signature" of how auditors interact with the BMD may differ from ordinary voters, enough to give clues to the malware about whether to cheat.[24] Therefore, one cannot simply multiply $(1 - p)^n$ and calculate a probability of detection.

While auditors might try to build an accurate model of voter behavior for live audits, that approach is doomed by privacy concerns and by the "curse of dimensionality": election officials would have to record every nuance of voter behavior (preferences

[24] For example, BMDs do "know" their own settings and other aspects of each voting session, so malware can use that information to target sessions that use the audio interface, increase the font size, use the sip-and-puff interface, set the language to something other than English, or take much longer than average to vote. (Voters who use those settings might be less likely to be believed if they report that the equipment altered their votes.) For parallel testing to have a good chance of detecting all outcome-changing problems, the tests must have a large chance of probing *every* combination of settings and voting patterns that includes enough ballots to change any contest result. It is not practical.

across contests; language settings, font settings, and other UI settings; timing, including speed of voting and hesitation; on-screen review; etc.) for million of voters to accurately approximate voter behavior.

There are many logistical problems with "live auditing." It would require additional voting machines (because testing requires additional capacity), staff, infrastructure, and other resources, *on election day* when professional staff is most stretched. One must be prepared to perform the audits at the busiest times of day, even that will cause lines of voters to lengthen, because otherwise the malware can simply cheat only at the busy times. Live auditing must be done in view of the voters (one cannot carry the voting machine into another room to do it), but some election officials are concerned that the creation of test ballots in the polling place could be perceived as a threat of ballot-box stuffing.

No state, to our knowledge has implemented parallel testing or live auditing of BMDs.

In any case, we can assess the contestability and defensibility of parallel testing.

With a sufficiently high rate of parallel testing, and a sufficiently sophisticated randomization of auditor behavior, it may be possible to make BMDs with parallel testing *contestable:* an audit could detect *and prove* mismarking of paper ballots.

But BMDs with parallel testing is not *defensible*. It will be extremely difficult for an election official to generate convincing public evidence that the audit *would have* detected mismarking, if mismarking were occurring. To generate that public evidence, the election official would have to reveal substantial detail about the parallel-testing protocol: how, exactly, the random selection of times to test is made; how, exactly, the random selection is made of what candidates to vote for in the tests. Revealing such details of the protocol allows the attacker to analyze the protocol for clues about how and when to cheat with less chance of detection.

Furthermore, parallel testing has a severe disadvantage in comparison with other contestable/defensible paper-ballot-based voting systems: If the auditors detect that the BMDs have mismarked a ballot—even once—the entire election must be invalidated, and a do-over election must be held. This is because the auditor will have detected evidence that the BMDs in this election have been systematically mismarking ballots for some proportion of *all* voters. No recount of the paper ballots can correct this.

In contrast, if optical scanners are hacked to cheat on hand-marked paper ballots,

the correct outcome can be calculated by a full hand recount of the paper ballots.[25]

Wallach also suggests, instead of parallel testing, the use of spoiled-ballot rates as a measure of BMD cheating. Suppose, when BMDs are not cheating the baseline rate of spoiled ballots (i.e., voters asking for a "do-over" of their BMD marked ballot) is 1%. Suppose the machines are cheating on 5% of the ballots, and 6% of voters notice this, and ask for a do-over. Then the spoiled ballot rate increases to 1.3%. The election administrator is supposed to act upon this discrepancy. But the only meaningful action the administrator could take is to invalidate the entire election, and call for a do-over election. This is impractical.

Moreover, the underlying "natural" rate of spoilage will not be known exactly, and will vary from election to election, even if the machines function flawlessly. The natural rate might depend on the number of contests on the ballot, the complexity of voting rules (e.g., IRV versus plurality), ballot layout, and many other factors. For any rule, there will be a tradeoff between false alarms and failures to detect problems.

To continue the previous hypothetical, suppose that spoiled ballots follow a Poisson distribution (there is no reason to think that they do). Imagine that the theoretical rate is known to be 1% if the BMDs function correctly, and known to be 1.3% if the BMDs malfunction. How many votes must be cast for it to be possible to limit the chance of a false alarm to 1%, while ensuring a 99% chance of detecting a real problem? The answer is 28,300 votes. If turnout is roughly 50%, jurisdictions (or contests) with fewer than 60,000 voters could not in principle limit the chance of false positives and of false negatives to 1%—even under these optimistic assumptions and simplifications. Twenty-three of California's 58 counties have fewer than 60,000 registered voters.

7 Other tradeoffs, BMDs versus hand-marked opscan

Supporters of ballot-marking devices advance several other arguments for their use.

- **Mark legibility.** A common argument is that a properly functioning BMD will generate clean, error-free, unambiguous marks, while hand-marked paper ballots may contain mistakes and stray marks that make it impossible to discern a voter's intent. However appealing this argument seems at first blush, the data are not nearly so compelling. Experience with statewide recounts in Minnesota

[25] Provided, of course, that secure chain of custody of the ballot boxes can be demonstrated.

and elsewhere suggest that truly ambiguous handmade marks are very rare.[26] For instance, 2.9 million hand-marked ballots were cast in the 2008 Minnesota race between Al Franken and Norm Coleman for the U.S. Senate. In a manual recount, between 99.95% and 99.99% of ballots were unambiguously marked.[27] [28] In addition, usability studies of hand-marked bubble ballots—the kind in most common use in U.S. elections—indicate a *voter* error rate of 0.6%, much lower than the 2.5–3.7% error rate for machine-marked ballots [17].[29] Thus, mark legibility is not a good reason to adopt BMDs for all voters.

- **Undervotes, overvotes.** Another argument offered for BMDs is that the machines can alert voters to undervotes and prevent overvotes. That is true, but modern PCOS systems can also alert a voter to overvotes and undervotes, allowing a voter to eject the ballot and correct it.
- **Bad ballot design.** Ill-designed paper ballots, just like ill-designed touchscreen interfaces, may lead to unintentional undervotes [25]. For instance, the 2006 Sarasota, Florida, touchscreen ballot was badly designed. The 2018 Broward County, Florida, opscan ballot was badly designed: it violated three separate guidelines from the EAC's 2007 publication, "Effective Designs for the Administration of Federal Elections, Section 3: Optical scan ballots." [40] In both of these cases (touchscreens in 2006, hand-marked optical-scan in 2018), undervote rates were high. The solution is to follow standard, published ballot-design guidelines and other best practices, both for touchscreens and for hand-marked ballots [3, 25].
- **Low-tech paper-ballot fraud.** All paper ballots, however they are marked, are vulnerable to *loss*, *ballot-box stuffing*, *alteration*, and *substitution* between the time they are cast and the time they are recounted. That's why it is so important

[26] States do need clear and complete regulations for interpreting voter marks.

[27] "During the recount, the Coleman and Franken campaigns initially challenged a total of 6,655 ballot-interpretation decisions made by the human recounters. The State Canvassing Board asked the campaigns to voluntarily withdraw all but their most serious challenges, and in the end approximately 1,325 challenges remained. That is, approximately 5 ballots in 10,000 were ambiguous enough that one side or the other felt like arguing about it. The State Canvassing Board, in the end, classified all but 248 of these ballots as votes for one candidate or another. That is, approximately 1 ballot in 10,000 was ambiguous enough that the bipartisan recount board could not determine an intent to vote." [1] See also [26]

[28] We have found that some local election officials consider marks to be ambiguous if *machines* cannot read the marks. That is a different issue from *humans* being unable to interpret the marks. Errors in machine interpretation of voter intent can be dealt with by manual audits: if the reported outcome is wrong because machines misinterpreted handmade marks, a RLA has a known, large chance of correcting the outcome.

[29] Better designed user interfaces (UI) might reduce the error rate for machine-marked ballots below the historical rate for DREs; however, UI improvements cannot keep BMDs from printing something other than what the voter is shown on the screen.

to make sure that ballot boxes are always in multiple-person (preferably bipartisan) custody whenever they are handled, and that appropriate physical security measures are in place. Strong, verifiable chain-of-custody protections are essential.

Hand-marked paper ballots are vulnerable to alteration by anyone with a pen. Both hand-marked and BMD-marked paper ballots are vulnerable to substitution: anyone who has poorly supervised access to a legitimate BMD during election day can create fraudulent ballots, not necessarily to deposit them in the ballot box immediately (in case the ballot box is well supervised on election day) but with the hope of substituting it later in the chain of custody.[30]

All those attacks (on hand-marked and on BMD-marked paper ballots) are fairly low-tech. There are also higher-tech ways of producing ballots indistinguishable from BMD-marked ballots for substitution into the ballot box if there is inadequate chain-of-custody protection.

- **Accessible voting technology.** When hand-marked paper ballots are used with PCOS, there is (as required by law) also an accessible voting technology available in the polling place for voters unable to mark a paper ballot with a pen. This is typically a BMD or a DRE. When the accessible voting technology is not the same as what most voters vote on—when it is used by very few voters—it may happen that the accessible technology is ill-maintained or even (in some polling places) not even properly set up by pollworkers. This is a real problem. One proposed solution is to require all voters to use the same BMD or all-in-one technology. But the failure of some election officials to properly maintain their accessible equipment is not a good reason to adopt BMDs for *all* voters. Among other things, it would expose all voters to the security flaws described above.[31] Other advocates object to the idea that disabled voters must use a different method of marking ballots, arguing that their rights are thereby violated. Both HAVA and ADA require reasonable accommodations for voters with physical and cognitive impairments, but neither law requires that those accommodations must be used by all voters. To best enable and facilitate participation by all voters, each voter should be provided with a means of casting a vote best suited to their abilities.
- **Ballot printing costs.** Preprinted optical-scan ballots cost 20–50 cents each.[32]

[30] Some BMDs print a barcode indicating when and where the ballot was produced, but that does not prevent such a substitution attack against currently EAC-certified, commercially available BMDs. We understand that systems under development might make ballot-substitution attacks against BMDs more difficult.

[31] Also, some accessibility advocates argue that requiring disabled voters to use BMDs compromises their privacy since hand-marked ballots are easily distinguishable from machine marked ballots. That issue can be addressed without BMDs-for-all: Accessible BMDs are already available and in use that mark ballots with marks that cannot easily be distinguished from hand-marked ballots.

[32] Single-sheet (one- or two-side) ballots cost 20-28 cents; double-sheet ballots needed for elections

Blank cards for BMDs cost up to 15 cents each, depending on the make and model of BMD.[33] But optical-scan ballots must be preprinted for as many voters as *might* show up, whereas blank BMD cards are consumed in proportion to how many voters *do* show up. The Open Source Election Technology Institute (OSET) conducted an independent study of total life cycle costs[34] for hand-marked paper ballots and BMDs in conjunction with the 2019 Georgia legislative debate regarding BMDs [27]. OSET concluded that, even in the most optimistic (i.e., lowest cost) scenario for BMDs and the most pessimistic (i.e, highest cost) scenario for hand-marked paper ballots and ballot-on-demand (BOD) printers—which can print unmarked ballots as needed—the total lifecycle costs for BMDs would be higher than the corresponding costs for hand-marked paper ballots.[35]

- **Vote centers.** To run a vote center that serves many election districts with different ballot styles, one must be able to provide each voter a ballot containing the contests that voter is eligible to vote in, possibly in a number of different languages. This is easy with BMDs, which can be programmed with all the appropriate ballot definitions. With preprinted optical-scan ballots, the PCOS can be programmed to *accept* many different ballot styles, but the vote center must still maintain *inventory* of many different ballots. BOD printers are another economical alternative for vote centers.[36]
- **Paper/storage.** BMDs that print summary cards rather than full-face ballots can save paper and storage space. However, many BMDs print full-face ballots—so they do not save storage—while many BMDs that print summary cards (which could save storage) use thermal printers and paper that is fragile and can fade in a few months.[37]

---

with many contests cost up to 50 cents.

[33] Ballot cards for ES&S ExpressVote cost about 15 cents. New Hampshire's (One4All / Prime III) BMDs used by sight-impaired voters use plain paper that is less expensive.

[34] They include not only the cost of acquiring and implementing systems but also the ongoing licensing, logistics, and operating (purchasing paper stock, printing, and inventory management) costs.

[35] BOD printers currently on the market arguably are best suited for vote centers, but less expensive options suited for polling places could be developed. Indeed, BMDs that print full-face ballots could be re-purposed as BOD printers for polling place use, with modest changes to the programming.

[36] Ballot-on-demand printers *may* require maintenance such as replacement of toner cartridges. This is readily accomplished at a vote center with a professional staff. Ballot-on-demand printers may be a less attractive option for many small precincts on election day, where there is no professional staff—but on the other hand, they are less necessary, since far fewer ballot styles will be needed in any one precinct.

[37] The California Top-To-Bottom Review (TTBR) of voting systems found that thermal paper can also be covertly spoiled wholesale using common household chemicals https://votingsystems.cdn.sos.ca.gov/oversight/ttbr/red-diebold.pdf, last visited 8 April 2019. The fact that thermal paper printing can fade or deteriorate rapidly might mean it does not satisfy the federal requirement to preserve voting materials for 22 months. http://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title52-section20701&num=0&edition=prelim, last visited 8

Advocates of hand-marked paper ballot systems advance these additional arguments.

- **Cost.** Using BMDs for all voters substantially increases the cost of acquiring, configuring, and maintaining the voting system. One PCOS can serve 1200 voters in a day, while one BMD can serve only about 260 [34]—though both these numbers vary greatly depending on the length of the ballot and the length of the day. OSET analyzed the relative costs of acquiring BMDs for Georgia's nearly seven million registered voters versus a system of hand-marked paper ballots, scanners, and BOD printers [27]. A BMD solution for Georgia would cost taxpayers between 3 and 5 times more than a system based on hand-marked paper ballots. Open-source systems might eventually shift the economics, but current commercial universal-use BMD systems are more expensive than systems that use hand-marked paper ballots for most voters.
- **Mechanical reliability and capacity.** Pens are likely to have less downtime than BMDs. It is easy and inexpensive to get more pens and privacy screens when additional capacity is needed. If a precinct-count scanner goes down, people can still mark ballots with a pen; if the BMD goes down, voting stops. Thermal printers used in DREs with VVPAT are prone to jams; those in BMDs might have similar flaws.

These secondary pros and cons of BMDs do not outweigh the primary security and accuracy concern: BMDs, if hacked or erroneously programmed, can change votes in a way that is not correctable. BMD voting systems are not contestable or defensible. Audits that rely on BMD printout cannot make up for this defect in the paper trail: they cannot reliably detect or correct problems that altered election outcomes.

Barcodes

A controversial feature of some BMDs allows them to print 1-dimensional or 2-dimensional barcodes on the paper ballots. A 1-dimensional barcode resembles the pattern of vertical lines used to identify products by their universal product codes. A 2-dimensional barcode or QR code is a rectangular area covered in coded image *modules* that encode more complex patterns and information. BMDs print barcodes on the same paper ballot that contains human-readable ballot choices. Voters using BMDs are expected to verify the human-readable printing on the paper ballot card, but the presence of barcodes with human-readable text poses some significant problems.

April 2019.

- **Barcodes are not human readable.** The whole purpose of a paper ballot is to be able to recount (or audit) the *voters'* votes in a way independent of any (possibly hacked or buggy) computers. If the official vote on the ballot card is the barcode, then it is impossible for the voters to verify that the official vote they cast is the vote they expressed. Therefore, before a state even *considers* using BMDs that print barcodes (and we do not recommend doing so), the State must ensure by statute that recounts and audits are based *only* on the human-readable portion of the paper ballot. Even so, audits based on untrustworthy paper trails suffer from the verifiability the problems outlined above.
- **Ballot cards with barcodes contain two different votes.** Suppose a state does ensure by statute that recounts and audits are based on the human-readable portion of the paper ballot. Now a BMD-marked ballot card with both barcodes and human-readable text contains two different votes in each contest: the barcode (used for electronic tabulation), and the human-readable selection printout (official for audits and recounts). In few (if any) states has there even been a discussion of the legal issues raised when the official markings to be counted differ between the original count and a recount.
- **Barcodes pose technical risks.** Any coded input into a computer system—including wired network packets, WiFi, USB thumbdrives, *and barcodes*—pose the risk that the input-processing software can be vulnerable to attack via deliberately ill-formed input. Over the past two decades, many such vulnerabilities have been documented on *each* of these channels (including barcode readers) that, in the worst case, give the attacker complete control of a system.[38] If an attacker were able to compromise a BMD, the barcodes are an attack vector for the attacker to take over an optical scanner (PCOS or CCOS), too. Since it is good practice to close down all such unneeded attack vectors into PCOS or CCOS voting machines (e.g., don't connect your PCOS to the Internet!), it is also good practice to avoid unnecessary attack channels such as barcodes.

8 Insecurity of All-in-One BMDs

Some voting machines incorporate a BMD interface, printer, and optical scanner into the same cabinet. Other DRE+VVPAT voting machines incorporate ballot-marking, tabulation, and paper-printout retention, but without scanning. These are often called

[38] An example of a barcode attack is based on the fact that many commercial barcode-scanner components (which system integrators use to build cash registers or voting machines) treat the barcode scanner using the same operating-system interface as if it were a keyboard device; and then some operating systems allow "keyboard escapes" or "keyboard function keys" to perform unexpected operations.

"all-in-one" voting machines. To use an all-in-one machine, the voter makes choices on a touchscreen or through a different accessible interface. When the selections are complete, the BMD prints the completed ballot for the voter to review and verify, before depositing the ballot in a ballot box attached to the machine.

Such machines are especially unsafe: like any BMD described in Section 3 they are not contestable or defensible, but in addition, if hacked they can print votes onto the ballot *after* the voter last inspects the ballot.

- The ES&S ExpressVote (in all-in-one mode) allows the voter to mark a ballot by touchscreen or audio interface, then prints a paper ballot card and ejects it from a slot. The voter has the opportunity to review the ballot, then the voter redeposits the ballot into the same slot, where it is scanned and deposited into a ballot box.
- The ES&S ExpressVoteXL allows the voter to mark a ballot by touchscreen or audio interface, then prints a paper ballot and displays it under glass. The voter has the opportunity to review the ballot, then the voter touches the screen to indicate "OK," and the machine pulls paper ballot up (still under glass) and into the integrated ballot box.
- The Dominion ImageCast Evolution (ICE) allows the voter to deposit a hand-marked paper ballot, which it scans and drops into the attached ballot box. *Or*, a voter can use a touchscreen or audio interface to direct the marking of a paper ballot, which the voting machine ejects through a slot for review; then the voter redeposits the ballot into the slot, where it is scanned and dropped into the ballot box.

In all three of these machines, the ballot-marking printer is in the same paper path as the mechanism to deposit marked ballots into an attached ballot box. This opens up a very serious security vulnerability: the voting machine can mark the paper ballot (to add votes or spoil already-cast votes) after the last time the voter sees the paper, and then deposit that marked ballot into the ballot box without the possibility of detection.

Vote-stealing software could easily be constructed that looks for *undervotes* on the ballot, and marks those unvoted spaces for the candidate of the hacker's choice. This is very straightforward to do on optical-scan bubble ballots (as on the Dominion ICE) where undervotes are indicated by no mark at all. On machines such as the ExpressVote and ExpressVoteXL, the normal software indicates an undervote with the words NO SELECTION MADE on the ballot summary card. Hacked software could simply leave a blank space there (most voters wouldn't notice the difference), and then fill in that space and add a matching bar code after the voter has clicked "cast this ballot."

An even worse feature of the ES&S ExpressVote and the Dominion ICE is the *auto-*

cast configuration setting (in the manufacturer's standard software) that allows the voter to indicate, "don't eject the ballot for my review, just print it and cast it without me looking at it." If fraudulent software were installed in the ExpressVote, it could change *all* the votes of any voter who selected this option, because the voting machine software would know *in advance of printing* that the voter had waived the opportunity to inspect the printed ballot. We call this auto-cast feature "permission to cheat" [4].

Regarding these all-in-one machines, we conclude:

- Any machine with ballot printing in the same paper path with ballot deposit is not *software independent*; it is *not* the case that "an error or fault in the voting system software or hardware cannot cause an undetectable change in election results." Therefore such all-in-one machines do not comply with the VVSG 2.0 (the Election Assistance Commission's Voluntary Voting Systems Guidelines). Such machines are not contestable or defensible, either.
- All-in-one machines on which all voters use the BMD interface to mark their ballots (such as the ExpressVote and ExpressVoteXL) *also* suffer from the same serious problem as ordinary BMDs: most voters do not review their ballots effectively, and elections on these machines are not contestable or defensible.
- The auto-cast option for a voter to allow the paper ballot to be cast without human inspection is particularly dangerous, and states must insist that vendors disable or eliminate this mode from the software. However, even disabling the auto-cast feature does not eliminate the risk of undetected vote manipulation.

Remark. The Dominion ImageCast Precinct ICP320 is a precinct-count optical scanner (PCOS) that also contains an audio+buttons ballot-marking interface for disabled voters. This machine can be configured to cast electronic-only ballots from the BMD interface, or an external printer can be attached to print paper optical-scan ballots from the BMD interface. When the external printer is used, that printer's paper path is *not* connected to the scanner+ballot-box paper path (a person must take the ballot from the printer and deposit it into the scanner slot). Therefore this machine is as safe to use as any PCOS with a separate external BMD.

9 Conclusion

Ballot-Marking Devices produce ballots that do not necessarily record the vote expressed by the voter when they enter their selections on the touchscreen: hacking, bugs, and configuration errors can cause the BMDs to print votes that differ from what the

voter entered and verified electronically. Because outcome-changing errors in BMD printout do not produce public evidence, BMD systems are not *contestable*. Because there is no way to generate convincing public evidence that reported outcomes are correct despite any BMD malfunctions that might have occurred, BMD systems are not *defensible*. Therefore, BMDs should not be used by voters who can hand mark paper ballots.

All-in-one voting machines, which combine ballot-marking and ballot-box-deposit into the same paper path, are even worse. They have all the disadvantages of BMDs (they are not contestable or defensible), and they can mark the ballot after the voter has inspected it. Therefore they are not even *software independent*, and should not be used by those voters who are capable of marking, handling, and visually inspecting a paper ballot.

When computers are used to record votes, the original transaction (the voter's expression of the votes) is not documented in a verifiable way.[39] When pen-and-paper is used to record the vote, the original expression of the vote *is* documented in a verifiable way (if demonstrably secure chain of custody of the paper ballots is maintained). Audits of elections conducted with hand-marked paper ballots, counted by optical scanners, can ensure that reported election outcomes are correct. Audits of elections conducted with BMDs *cannot* ensure that reported outcomes are correct.

References

[1] A.W. Appel. Optical-scan voting extremely accurate in Minnesota. *Freedom to Tinker*, January 2009. https://freedom-to-tinker.com/2009/01/21/optical-scan-voting-extremely-accurate-minnesota/.

[2] A.W. Appel. End-to-end verifiable elections. *Freedom to Tinker*, November 2018. https://freedom-to-tinker.com/2018/11/05/end-to-end-verifiable-elections/.

[3] A.W. Appel. Florida is the Florida of ballot-design mistakes. *Freedom to Tinker*, November 2018. https://freedom-to-tinker.com/2018/11/14/florida-is-the-florida-of-ballot-design-mistakes/.

[39] It is conceivable that cryptographic protocols like those used in E2E-V systems could be used to create BMD-based systems that are contestable and defensible, but no such system exists, nor, to our knowledge, has such a design been worked out in principle. Existing E2E-V systems that use a computer to print (encrypted) selections are neither contestable nor defensible, as explained in Section 1.

[4] A.W. Appel. Serious design flaw in ESS ExpressVote touchscreen: "permission to cheat". *Freedom to Tinker*, September 2018. https://freedom-to-tinker.com/2018/09/14/serious-design-flaw-in-ess-expressvote-touchscreen-permission-to-cheat/.

[5] J. Benaloh, M. Byrne, B. Eakin, P. Kortum, N. McBurnett, O. Pereira, P.B. Stark, , and D.S. Wallach. Star-vote: A secure, transparent, auditable, and reliable voting system. *JETS: USENIX Journal of Election Technology and Systems*, 1:18–37, 2013.

[6] J. Benaloh, D. Jones, E. Lazarus, M. Lindeman, and P.B. Stark. SOBA: Secrecy-preserving observable ballot-level audits. In *Proceedings of the 2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX, 2011.

[7] Josh Benaloh, Ronald L. Rivest, Peter Y. A. Ryan, Philip B. Stark, Vanessa Teague, and Poorvi L. Vora. End-to-end verifiability. *CoRR*, abs/1504.03778, 2015.

[8] Matthew Bernhard, Allison McDonald, Henry Meng, Jensen Hwa, Nakul Bajaj, Kevin Chang, and J. Alex Halderman. Can voters detect malicious manipulation of ballot marking devices? In *41st IEEE Symposium on Security and Privacy*, page (to appear). IEEE, 2020.

[9] R. K. Bothwell, K.A. Deffenbacher, and J.C. Brigham. Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, 72:691–695, 1987.

[10] D. Chaum, A. Essex, R.T. Carback III, J. Clark, S. Popoveniuc, A.T. Sherman, and P. Vora. Scantegrity: End-to-end voter verifiable optical-scan voting. *IEEE Security & Privacy*, 6:40–46, 2008.

[11] Election Assistance Commission. Voluntary voting systems guidelines 2.0, September 2017. https://www.eac.gov/assets/1/6/TGDC_Recommended_VVSG2.0_P_Gs.pdf.

[12] Moritz Contag, Guo Li, Andre Pawlowski, Felix Domke, Kirill Levchenko, Thorsten Holz, and Stefan Savage. How they did it: An analysis of emission defeat devices in modern automobiles. In *2017 IEEE Symposium on Security and Privacy*, pages 231–250. IEEE, 2017.

[13] K. Deffenbacher. Eyewitness accuracy and confidence: Can we infer anything about their relation? *Law and Human Behavior*, 4:243–260, 1980.

[14] R. DeMillo, R. Kadel, and M. Marks. What voters are asked to verify affects ballot verification: A quantitative analysis of voters' memories of their ballots, November 2018. https://ssrn.com/abstract=3292208.

[15] S.L. Desmarais, T.L. Nicholls, J. D. Read, and J. Brink. Confidence and accuracy in assessments of short-term risks presented by forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology*, 21(1):1–22, 2010.

[16] D. Dunning, D.W. Griffin, J.D. Milojkovic, and L. Ross. The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, 58:568–581, 1990.

[17] S.P. Everett. *The Usability of Electronic Voting Machines and How Votes Can Be Changed Without Detection*. PhD thesis, Rice University, 2007.

[18] A.J. Feldman, J.A. Halderman, and E.W. Felten. Security analysis of the Diebold AccuVote-TS voting machine. In *2007 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT 2007)*, August 2007.

[19] Verified Voting Foundation. The verifier – polling place equipment – november 2018, November 2018. https://www.verifiedvoting.org/verifier/.

[20] P. Johansson, L. Hall, and S. Sikstrom. From change blindness to choice blindness. *Psychologia*, 51:142–155, 2008.

[21] D. Kahnemann. *Thinking, fast and slow*. Farrar, Straus and Giroux, 2011.

[22] S. J. Lewis, O. Pereira, and V. Teague. Ceci n'est pas une preuve: The use of trapdoor commitments in Bayer-Groth proofs and the implications for the verifiabilty of the Scytl-SwissPost Internet voting system, 2019. https://people.eng.unimelb.edu.au/vjteague/UniversalVerifiabilitySwissPost.pdf.

[23] M. Lindeman and P.B. Stark. A gentle introduction to risk-limiting audits. *IEEE Security and Privacy*, 10:42–49, 2012.

[24] National Academies of Sciences, Engineering, and Medicine. *Securing the Vote: Protecting American Democracy*. The National Academies Press, Washington, DC, September 2018.

[25] L. Norden, M. Chen, D. Kimball, and W. Quesenbery. Better Ballots, 2008. Brennan Center for Justice, http://www.brennancenter.org/publication/better-ballots.

[26] Office of the Minnesota Secretary of State. Minnesota's historic 2008 election, 2009. https://www.sos.state.mn.us/media/3078/minnesotas-historic-2008-election.pdf.

[27] E. Perez. Georgia state election technology acquisition: A reality check. OSET Institute Briefing, March 2019. https://trustthevote.org/wp-content/uploads/2019/03/06Mar19-OSETBriefing_GeorgiaSystemsCostAnalysis.pdf.

[28] K. Rayner and M.S. Castelhano. Eye movements during reading, scene perception, and visual search, 2009. Q J Experimental Psychology, 2009, August 62(8), 1457-1506.

[29] J. Reason. *Human Error (20th Printing)*. Cambridge University Press, New York, 2009.

[30] R.L. Rivest and J.P. Wack. On the notion of software independence in voting systems, July 2006. http://vote.nist.gov/SI-in-voting.pdf.

[31] Ronald L Rivest. On the notion of 'software independence' in voting systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1881):3759–3767, 2008.

[32] Ronald L Rivest and Madars Virza. Software independence revisited. In *Real-World Electronic Voting*, pages 19–34. Auerbach Publications, 2016.

[33] P.Y.A. Ryan, D. Bismark amnd J. Heather, and S. Schneiderand Z. Xia. The prêt à voter verifiable election system. *IEEE Transactions on Information Forensics and Security*, 4:662–673, 2009.

[34] Election Systems and Software. State of Georgia Electronic Request for Information New Voting System Event Number: 47800-SOS0000035, 2018. http://sos.ga.gov/admin/files/ESS%20RFI%20-%20Final%20-%20Redacted.pdf.

[35] P.B. Stark. Conservative statistical post-election audits. *Annals of Applied Statistics*, 2:550–581, 2008.

[36] P.B. Stark. Risk-limiting post-election audits: P-values from common probability inequalities. *IEEE Transactions on Information Forensics and Security*, 4:1005–1014, 2009.

[37] P.B. Stark. An introduction to risk-limiting audits and evidence-based elections, 2018. Testimony prepared for the California Little Hoover Commission, https://www.stat.berkeley.edu/~stark/Preprints/lhc18.pdf.

[38] P.B. Stark. There is no reliable way to detect hacked ballot-marking devices. https://arxiv.org/abs/1908.08144, 2019.

[39] P.B. Stark and D.A. Wagner. Evidence-based elections. *IEEE Security and Privacy*, 10:33–41, 2012.

[40] U. S. Election Assistance Commission. Effective designs for the administration of federal elections, June 2007. https://www.eac.gov/assets/1/1/EAC_Effective_Election_Design.pdf.

[41] Dan S. Wallach. On the security of ballot marking devices, December 2019.

[42] J.T. Wixted and G.L. Wells. The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, 2017.