Some Consequences of Paper Fingerprinting for Elections

Joseph A. Calandrino[*], William Clarkson[*], and Edward W. Felten[*,†]

[*]Center for Information Technology Policy and Dept. of Computer Science, Princeton University
[†]Woodrow Wilson School of Public and International Affairs, Princeton University
{jcalandr,wclarkso,felten}@cs.princeton.edu

Abstract

Recent research has demonstrated that individual pieces of paper can be fingerprinted and reidentified later at low cost, using commodity scanners. We consider the consequences of this fact for electronic voting. The most obvious consequence is negative: the ability to fingerprint paper endangers the secrecy of ballots in any system that keeps paper records of individual ballots, including standard optical scan and DRE-VVPAT systems. We characterize the resulting risks and discuss when and how they can be mitigated. Less obviously, the ability to fingerprint paper can also have positive consequences, by enabling certain new kinds of post-election audit procedures, both to compare electronic records to the corresponding paper records and to detect the use of forged ballot stock. Paper reidentification presents new challenges for election officials, but careful consideration of its implications now may allow us to preserve ballot secrecy and strengthen election integrity.

1 Introduction

Paper records play a pivotal role in many voting systems. Paper is cheap, familiar, and reliable; and paper records can be read and written by people and machines. As a result of paper's positive properties, the voting systems most widely recommended by election security experts rely on keeping paper records of each ballot.

Even when using paper, however, care must be taken to ensure election integrity and ballot secrecy. Over the past several years, advances in algorithms and computer hardware have enabled paper fingerprinting—unique identification of individual sheets of paper based on physical characteristics [13, 4, 18, 8]. State-of-the-art techniques can identify a piece of paper without any need to mark or alter it in advance and require no more than a commodity desktop scanner and personal computer. The ability to uniquely identify a sheet of paper has both negative and positive implications for voting systems that rely on paper ballots. In certain scenarios, tracking of individual paper ballots could undermine ballot secrecy, while in others it can enable more efficient auditing techniques. In this paper, we discuss both the good and the bad, exploring various threat scenarios and proposing an auditing scheme that relies on fingerprinting individual ballots. Through vigilance and careful procedures, election officials may mitigate many threats posed by paper fingerprinting while harnessing its benefits.

Traditional paper-based voting systems, optical scan voting systems, and DRE-VVPAT systems[1] result in paper ballots containing the voters' choices for various contests. Suppose that someone has access to the paper ballots (or scans of the ballots) before and after an election. If this person can identify the paper ballot you will use to vote, then when election day concludes, that person can reidentify and recover the paper ballot containing your votes.[2]

Recent advances show that it is possible to identify sheets of paper based on unique physical characteristics. These characteristics are often imperceptible to the human eye, yet accurately measured by a machine. The most advanced systems measure slight variations in color or 3D surface texture of paper, requiring only a commodity desktop scanner and custom software. This is accomplished without modifying the paper in any way. Fingerprinting of paper ballots presents additional privacy challenges that must be addressed by election officials to ensure ballot secrecy. Fortunately, many attacks based on fingerprinting are non-trivial; they require access to the paper ballots and certain equipment at particular times. We survey the risks that fingerprinting poses based on details of common voting systems and the malicious party's

[1] Direct recording electronic voting machines with a voter-verified (or voter-verifiable) paper trail—a computerized voting system producing redundant paper ballot records that each voter may verify and approve.

[2] Some fingerprinting techniques can reidentify paper even if the voting process results in markings or creases.

level of access. Based on common themes, we provide suggestions for election officials to minimize these risks.

The ability to uniquely reidentify paper ballots also has implications for election auditing. When performing an audit, we may select individual ballots for review based on fingerprints, enabling efficient ballot-based auditing. A ballot's fingerprint often depends on physical characteristics of the paper that are indistinguishable to the human eye, yet still identifiable by a machine. These fingerprints effectively serve as serial numbers on the ballots without posing the same privacy risks as serial numbers. Paper fingerprints may also help uncover attacks that are problematic for some auditing schemes, including attacks that rely on ballot box stuffing. These additional checks can potentially result in greater election integrity. Because the security of the auditing proposals relies on the fingerprinting process, we briefly discuss the security of paper fingerprinting.

The remainder of this paper is organized as follows. Section 2 provides background information on the topic of paper fingerprinting as it is relates to fingerprinting paper ballots. Section 3 describes new threats posed by paper fingerprinting as well as choices that may mitigate these threats. Section 4 proposes new auditing techniques that make use of paper fingerprinting. Section 5 briefly discusses the impact of paper fingerprinting on an end-to-end voting scheme. Finally, Section 6 concludes.

2 Fingerprinting Background

In the past few years, several systems capable of identifying pieces of paper based on difficult-to-reproduce physical characteristics have been developed [13, 4, 18, 8]. These fingerprinting systems can rely on any distinguishing characteristics of paper. For example, when viewed up close, the surface of a sheet of paper is not perfectly flat, but actually a tangled matte of paper fibers. This non-uniformity in a paper's surface texture (the 3D shape of a paper's surface) is created during the manufacturing process and is unique to each sheet of paper. The FiberFingerprint system of Metois et al. first introduced the notion of using surface texture to uniquely identify a document [13]. FiberFingerprint uses a custom device to measure "inhomogeneities in the substrate" of a document, from which a unique identifier is derived. Laser Surface Authentication is a more advanced technique that measures a document's surface texture using a high-powered laser microscope [4]. Other methods, such as Print Signatures, can identify sheets of paper based on different unique characteristics, such as the random ink splatter that occurs around the edges of any features printed on it [18].

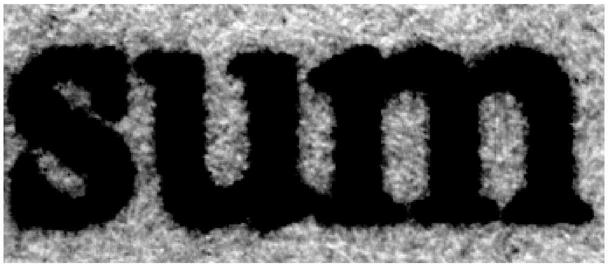
Recent work by Clarkson et al. describes another method for identifying sheets of paper based on surface texture [8]. Clarkson's method uses only a commodity scanner to reconstruct a paper's surface texture, from which a secure and robust fingerprint is derived. This technique requires no costly or specialized equipment, relying only on an unmodified commodity desktop scanner and custom software, and produces fingerprints that are verifiable using any appropriate-resolution scanner. In addition, the process does not require modification of the paper in any way, leaving no evidence that it was fingerprinted.

During an election, paper ballots may be treated harshly. These ballots may be creased or folded, and they are modified by markings indicating voter choices. Any paper identification system must be robust to harsh treatment and moderate levels of marking. We focus on the methods developed in [8] because this system utilizes inexpensive, widely available equipment and can reidentify a document even when a sheet of paper is handled harshly or modified by being printed or scribbled on. We briefly review two variants of the system from [8], one requiring only a single scan and another more secure version that requires multiple scans. We then discuss a method for protecting the privacy of paper ballot fingerprints. For more details, please refer to [8].

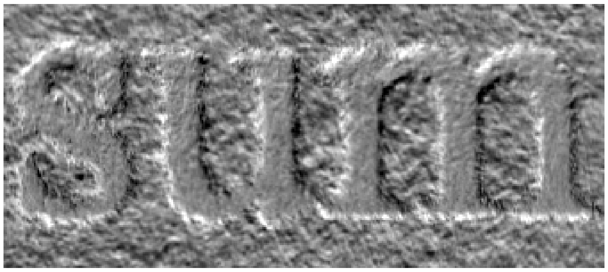
Single Scan. This scheme identifies a sheet of paper based on slight variations in its color. After scanning a sheet of paper at a high resolution—at least 1200 DPI—and adjusting the contrast we can see slight variations in color due to surface texture. See Figure 1(a). A fingerprint for the sheet of paper is derived from these variations. However, the resulting fingerprint may not be highly secure against forgery, as high quality printers may be able to mimic the measured feature.

Multiple Scans. This more secure version reconstructs the surface texture of a sheet of paper by taking multiple scans at different orientations. This is accomplished without changing the paper in any way, using only an unmodified commodity scanner and custom software [8]. A sheet's surface texture is estimated by scanning the document at four orientations and combining the data inferred from each scan. The level of detail is dramatic, see Figure 1(b). After measuring the surface texture, a concise fingerprint for the document is generated. The fingerprint under this variant is secure to forgery, as the surface texture of paper is difficult to duplicate or precisely modify. The resulting document fingerprint is also robust, allowing successful re-identification even if the document is creased, scribbled on, soaked in water, or printed on [8].

Protecting Ballot Fingerprints. Regardless of which variant is used to generate a ballot's fingerprint, knowl-



(a) A contrast adjusted 1200 DPI scan of a sheet of paper. In the background, we see slight color variations, which are used to identify individual sheets of paper. Actual size: “sum”.



(b) Image depicting the surface texture as measured by two 1200 DPI scans. After only two scans, the geometry of the paper surface is clearly visible. The embossing effect of the text is actually the wax ink sitting on top of the paper surface.

edge of the raw fingerprint could aid an adversary in generating a counterfeit sheet. If ballot fingerprints are published, then the privacy of each fingerprint must be protected, using a secure sketch [10] or some other method.

A secure sketch protects the fingerprint from accidental disclosure and only allows verification when provided with a document close to the original. Intuitively a secure sketch is a fuzzy hash function, with similar (nearby according to an appropriate metric) inputs producing identical outputs. Similar to cryptographically secure hash functions, given an output it is infeasible to find a input that produces identical output. The verification process of a fingerprint compares the fuzzy hash of a ballot's fingerprint against a list of candidates. This enables the secure verification of ballot fingerprints. As Dodis et al. demonstrate, a ballot's secure sketch does not reveal the underlying fingerprint, and thus the ballot's surface texture [10]. In addition, the methods for fingerprinting ballots described in [8] will not reveal the votes that these ballots contain. This permits the publication of a list of secure sketches of ballot fingerprints, without revealing the fingerprints themselves. This list can later be used to verify that a ballot is legitimate (among other things) by comparing it's fingerprint to those on the list.

3 New Threats

Although the ability to fingerprint and reidentify paper ballots poses a serious threat to voter privacy, officials may take steps to mitigate these threats. In this section, we detail several threats to voting systems utilizing paper ballots and discuss possible countermeasures.

Even without fingerprinting techniques, a number of methods exist for making paper ballots unique and potentially identifiable. Voters may choose an unusual write-in candidate or a unique combination of candidates to create a distinctive ballot. Given enough races, a pseudorandom selection of choices would with high probability create a distinctive ballot that could be later identified. Alternatively, poll workers may mark ballots with invisible ink or lightly crease the corner of a ballot. The paper fingerprinting system in [8] differs from past ballot tracing techniques (including invisible ink) because there is no way to detect, even by close inspection, whether a ballot is traceable. Every ballot is inherently traceable.

We consider only threats created or made easier by fingerprinting. For example, suppose that an optical scan system stores a fingerprint or high resolution scan of every ballot in the order they are cast. Given this information, an observer that watches when voters submit their ballots and can later examine the ballots for fingerprints could reidentify a voter's paper ballot, thus revealing the voters' choices. An easier, equally devastating attack exists without fingerprinting, however: the optical scan machine could simply store all votes cast in order. Therefore, this threat is not considered.

The threats we consider are those in which someone other than a voter can learn the voter's choices for various contests. This may be with the consent of the voter. To sell her vote, a voter may provide the fingerprint of her paper ballot to a purchaser. If paper ballots are revealed on or after election day, the purchaser can reidentify the appropriate ballot and verify the choices. Alternatively, someone may want the ability to uncover unconsenting voters' choices—whether due to curiosity or a desire to coerce voters—by fingerprinting blank paper ballots in advance and later associating voters with completed ballots. Anyone from voters and election officials to paper mill workers may be a participant in these threat scenarios. We use the term adversary to refer to any party that seeks to undermine ballot secrecy.

This reveals two general classes of attacks relating to paper identification. The first occurs when an adversary is able to fingerprint a significant portion of the ballot stock prior to an election and later associate voters with particular ballots. The second occurs when an individual voter scans her own ballot, either voluntarily or under coercion. We now discuss threats for various voting systems under each attack model.

3.1 Ballot Stock Fingerprinting

In this attack model, an adversary with access to the ballot stock is able to fingerprint or make high-quality scans of a significant portion of the ballots. By combining the ballot fingerprints with information about the order of voters an adversary may be able to undermine ballot secrecy. We discuss this attack for various voting systems.

DRE-VVPAT with Paper Spool. Paper fingerprinting poses a serious threat to DRE-VVPAT systems using printers with paper spools. We do not consider continuous spool-to-spool DRE-VVPAT systems, as they fundamentally fail to protect the secret ballot [12]. In cut-and-drop VVPAT systems, the record tape is used in a fixed order. As each voter casts his ballot, the tape is cut, dropping the segments into a box. Suppose that an adversary has unmonitored access to this paper spool prior to election day. The adversary can unroll the spool and repeatedly fingerprint short segments of the paper tape in order, storing the fingerprints in that order. He can then re-roll the spool and return it. Because the DRE will use the paper tape in order, the order of the segment fingerprints will correspond to the order that ballots are cast on the DRE. Although some segments may be destroyed when the ballots are separated, use of short enough segments can ensure that at least one segment remains intact per paper ballot, allowing complete re-ordering. For this attack to succeed, an adversary would need the ability to scan paper tape segments prior to the election, to observe the order that some or all voters enter the voting booth, and to fingerprint some or all of the resulting paper ballots later.

It is important to note that an adversary need not observe all voters or fingerprint all ballots before and after the election. Suppose that an adversary can reidentify her own ballot without scanning it, perhaps by casting a distinctive ballot, creating a point reference with respect to future ballots. If you enter the voting booth immediately after the adversary, she knows that your ballot will contain the segments immediately following her ballot's segments. Therefore, if the adversary can determine the fingerprint of her ballot after the election, she can guess the fingerprint on your ballot. Numerous similar correlation attacks are possible.

DRE-VVPAT with Paper Sheets. Using a DRE-VVPAT with standard, disconnected paper sheets does not mitigate all threats of a paper spool. Anyone that can fingerprint these sheets and has some knowledge of the order in which they will be loaded can mount attacks similar to those involving paper spools. In this case, an adversary would need the ability to scan some or all paper ballots prior to the election, to gain some knowledge of the order that these paper sheets will be loaded into the DRE, to observe the order that some or all voters enter the voting booth, and to fingerprint some or all of the resulting paper ballots later.

Paper-Based Voting. The issues with paper-based voting are similar to DRE-VVPAT with paper sheets. In this case, a voter may (potentially) have the ability to choose his own blank paper ballot, but the voter may be paid, coerced, or confused into making a non-random choice, such as taking the top ballot on the pile. Instead we recommend giving each voter an opportunity to re-randomize the pile of ballots, for example, by cutting the deck. Following this shuffling, an adversary would have greater difficulty inferring a relationship between future voters and their ballot fingerprints. To undermine a paper-based voting system, an adversary would need the ability to scan some or all paper ballots prior to the election, to gain some knowledge of the order of these paper sheets at the poll workers' table, to observe the order that some or all voters receive paper ballots (and, if voters can randomly choose or shuffle the ballots, to draw meaningful inferences in spite of this uncertainty), and to fingerprint some or all of the resulting paper ballots later.

Optical Scan Voting. In many ways, optical scan voting machines present a similar scenario to paper-based voting, but these machines also contain a scanner that may be capable of fingerprinting the ballots that they scan.[3] The optical scan machine may store the fingerprint and possibly votes on each ballot as part of its normal operations and publicly reveal this data later (Section 4 describes how this may be helpful for auditing). In this case, a malicious party would need the ability to scan some or all paper ballots prior to the election, to gain some knowledge of the order of these paper sheets, to observe the order that some or all voters receive paper ballots (and, if voters can randomly choose or shuffle the ballots, to draw meaningful inferences in spite of this uncertainty), and to fingerprint some or all of the resulting paper ballots later (or if the voting machine records the fingerprint-vote combinations electronically, to observe these values).[4]

[3] Even if infeasible with present machines, it may become feasible as low-cost scanners increase in resolution and gain additional capabilities.

[4] A reviewer notes an additional interesting attack. Assume that each paper ballot contains a stub that includes a serial number, and these stubs are removed for privacy reasons before a voter submits her ballot. Suppose that an adversary can scan ballots and associate fingerprints with each serial number prior to the election. If the adversary can observe the serial number of a ballot given to a voter, that adversary immediately knows the fingerprint of that voter's ballot. This attack emphasizes that, even if removed, serial numbers or other identifiable attributes on a ballot can threaten voter privacy.

Pre-Completed Ballots. Fingerprints can also enable dangerous coercion attacks utilizing pre-completed ballots. Imagine that an adversary wishes to coerce voters into choosing a particular candidate. The adversary can distribute pre-completed ballots to targeted individuals, recording the fingerprint of the ballot provided to each voter. If the adversary has access to the ballots after the election, she can use the fingerprints to reidentify the distributed ballots. This would allow the adversary to confirm that the provided ballots are in the ballot box and contain the "correct" votes.

3.1.1 Mitigation.

The various attack scenarios that we discuss each place certain requirements on an attacker. By making these requirements more difficult to achieve, we may reduce the feasibility of these threats.[5]

In all cases discussed in Section 3.1, an adversary must scan some or all paper ballots before voters cast those ballots. To prevent an adversary from producing scans, the paper sheets or rolls should be kept in a locked box whenever possible, and access to the paper should be monitored prior to election day. Election officials should make every effort to prevent the introduction of outside, rather than official, ballots into the ballot box. Scanners and computing devices should be kept away from the paper ballots.

Except in the case of DRE-VVPAT with paper spools, an adversary needs the ability to completely or partially learn the order of these paper sheets. This can be mitigated by shuffling the sheets immediately prior to election day. Depending on how well-shuffled the ballots are, the order information necessary to mount an attack may be destroyed. Shuffling is far from ideal, but it can make attacks harder.

These scenarios also rely on an adversary's ability to observe the order that voters obtain their paper ballots or the order that these voters enter the voting booth. Given the need for poll workers and observers to view the process, we unfortunately cannot eliminate these possibilities. To mitigate the threats, however, voters should have the ability (if reasonably possible) to shuffle or otherwise re-randomize the pile of paper ballots. This protects other voters too, as randomization of ballots increases an adversary's uncertainly in the relationship between voters and fingerprints.

Alternatively, cryptographic techniques may assist in ensuring properly shuffled ballots. For example, see [17], which considers a similar problem. At this time, we have reservations regarding the practicality and security of applying such techniques to this problem, but existing or future techniques may prove themselves to be efficient and secure if carefully applied to this unique scenario.

In many cases, an adversary requires the ability to scan certain ballots following the election. Except as necessary for auditing and other processes, scanners should not be allowed near the ballots following the election. In general, used ballots should be stored securely in a monitored location.

As described earlier, some optical scan machines may record fingerprint-vote combinations electronically, and the adversary may use this information to reidentify voters' ballots. These values should be stored and revealed only to the degree necessary to conduct the election and audit processes.

No countermeasure discussed in this section is perfect on its own. In addition, some countermeasures may inhibit other necessary goals, such as the ability to conduct a secure, efficient audit. Election officials may choose to focus their efforts on a limited number of feasible countermeasures to eliminate this threat. A cautious ballot shuffling process probably has the greatest potential to eliminate the threat of fingerprinting to ballot secrecy with minimal impact on the remainder of the election process.

3.2 Individual Ballot Fingerprinting

In this scenario, a voter reveals his votes to a third party, perhaps under coercion or to permit vote selling. Suppose that a coercer tells a voter to scan his paper ballot between receiving and submitting it and to return a fingerprint of the scan to the coercer. After the election, the coercer can verify fingerprints to identify the voter's ballot among the set of legitimate ballots. If the coercer does not find the ballot or if the ballot contains "incorrect" votes, the voter may face repercussions.

Similarly, a voter may sell her vote by providing a purchaser with a scan or fingerprint of her ballot. If a legitimate ballot exists with that fingerprint and contains the correct votes, the purchaser will pay the voter.

3.2.1 Mitigation

Attacks based on the ability of individual voters to fingerprint their ballots are difficult to prevent. If voters have the covert ability to measure inherent, unique physical properties of their own ballots, election officials are left with few methods of recourse. Our best option to mitigate this threat is disallowing scanners or similar devices near the ballots throughout the election process. While the prospect of a voter sneaking a scanner into the voting booth may seem far-fetched, handheld scanners,

[5] We do not consider mitigation techniques that seek to make paper ballots more difficult to fingerprint by modifying the ballots or using a different material. Even if possible, such solutions are likely both to be costly and to be vulnerable to future specialized attacks.

increasing-quality cell phone cameras, and other technological innovations are increasing the practicality of these attacks.

Arguably, voters have long had the ability to make their ballots stand out through unique write-in choices or combinations of votes. The threat of fingerprinting differs, however, because a fingerprinted ballot is reidentifiable even if no affirmative steps are taken to make the ballot unique.

4 Auditing Techniques

The popularity of paper ballots is partially a consequence of concerns regarding flaws and vulnerabilities in computerized voting systems. Software cannot change a paper ballot already in a ballot box, making voter-verified paper ballots a popular mitigation strategy. These paper ballots are only useful if someone verifies that they are consistent with the electronically tabulated outcome.

Rather than recount all paper ballots to verify the election results, we may examine some subset of these ballots to draw statistical inferences about the election's outcome. The most popular, widely used auditing approach is precinct-based auditing (e.g., [1, 3, 2, 15, 16]). With precinct-based auditing, officials and other parties randomly select some subset of election precincts. For the selected precincts, officials manually count the votes on all paper ballots and ensure that they match the electronic results.

Sampling at a finer level of granularity than precincts can allow for equally strong statistical inferences with fewer paper ballots manually reviewed. The finest level of granularity possible is ballot-based auditing, in which auditors select some subset of electronic ballots and ensure that they match their corresponding paper ballots (e.g., [6, 14, 11]). Ballot-based auditing presents a number of subtle challenges. For example, it can be difficult to ensure that ballots are selected at random without compromising ballot secrecy.

Calandrino et al. describe a machine-assisted auditing method that allows ballot-based auditing yet strives to preserve ballot secrecy [6]. Following the election, an auditing machine prints serial numbers on paper ballots and rescans the ballots, storing the serial numbers and votes electronically. If these votes sum to the initially reported electronic tally, auditors randomly select electronic ballots based on serial number and manually verify that they match the corresponding paper ballots. The scheme in [6] is able to detect discrepancies even if the auditing machine misbehaves. We propose a method that allows ballot-based auditing without printing serial numbers on ballots, instead relying on paper ballot fingerprints.

For our auditing method, we make several assumptions. First, we assume that precincts maintain a sign-in list that observers may monitor as voters enter and leave a polling place. As is standard practice, the sign-in list is made public after the election. This list allows anyone to determine an accurate count of the number of voters. We assume that every voter signing in casts a ballot. In practice, voters rarely sign in without casting a ballot, and we leave methods for ascertaining the number of ballots cast to future work (any issues with obtaining an accurate count affects all known auditing schemes, not just the ones in this paper). Our auditing scheme checks for discrepancies not only between the paper and electronic records but between those records and the totals from the sign-in list.

Many ballot-based auditing schemes assume that the set of paper ballots contains no added ballots. Section 4.2 describes serious additional threats that are possible if an adversary is able to add paper ballots to the ballot box—even if the set of correct paper ballots remains. These attacks can be difficult to detect, traditionally requiring an accurate count of the paper ballots. We discuss this threat and describe how to use fingerprints to detect added paper ballots more efficiently.

Auditors might also want the ability to verify that the set of paper ballots in the ballot box matches the ones delivered to the polling place before the election. This capability enables a number of possibilities. For example, officials could find the set of legitimate paper ballots in the event of ballot-box stuffing. Section 4.3 describes how to perform this check by fingerprinting ballots immediately prior to the election.

4.1 Fingerprint-Based Auditing

We first discuss a scheme for using fingerprints to detect mismatches from the electronic ballots to the corresponding paper ballots in the ballot box. Our process is designed such that, if at least B incorrect electronic ballots exist, we will find one or more discrepancies with probability greater than or equal to a desired confidence level c.[6] This scheme is primarily for demonstrative purposes, as it requires revelation of the fingerprint and full combination of votes for each ballot cast, potentially posing serious privacy concerns. This requirement undermines ballot secrecy if a voter can distinguish her ballot through her votes.

Throughout this section, we assume use of an optical scan voting machine. Our scheme works as follows.

[6] In practice, we want the audit process to determine whether we can be confident in the election's outcome even if we observe a small number of discrepancies. Although we use the more conventional goal of election auditing in this paper, our methods can extend to meet the more ambitious practical goal.

General Election

Issue	Totals		Fixed Sample Size	Varying Sample Size	Pct-Based Auditing
	# Votes	Margin	# Bal (Man)	# Bal (Man)	# Bal (Man)
U.S. Senate	2,370,445	0.39%	2,337	2,339	1,141,900
Const. Amnd.	2,328,224	14.12%	63	65	8,062
U.S. House	173,159	2.82%	325	327	62,469
U.S. House	212,079	19.19%	46	48	1,958
U.S. House	241,134	16.36%	54	56	6,120
U.S. House	235,280	11.88%	76	77	12,991
Delegate	14,963	5.75%	157	159	11,442
Average	796,469	8.11%	437	439	177,849

Table 1: Ballot-Based Auditing with Fingerprints (For Both Fixed and Varying Sample Size Methods) vs. Precinct-Based Auditing on 2006 Virginia General Election Data. (Note: these numbers are originally from [6], which describes post-election auditing methods that yield equivalent sample sizes to the methods in this paper.)

When a voter submits her paper ballot, the voting machine records both the ballot's fingerprint and a vector of the votes contained on that ballot. At the end of the election, officials immediately publish both these fingerprint-vote pairs and the voter sign-in sheet (as discussed earlier, the sign-in sheet should be public throughout election day). Therefore, any citizen will have the ability to confirm that the number of electronic ballots matches the number of signed-in voters and that the votes posted add up to the reported tallies. Our final step is to sample from the electronic ballots and ensure that matching paper ballots exist. Otherwise, a discrepancy exists. Note that switching a vote on a single ballot from Candidate A to Candidate B affects the margin between the candidates by two: Candidate A loses one vote and Candidate B gains one vote. Therefore, we seek to reject the hypothesis that $B = margin/2$ incorrect electronic ballots exist.

To sample from the electronic ballots, we make a list of the (precinct, fingerprint, vote vector) triples, one for each ballot, ordered lexicographically. We will sample items from this list and ensure that a ballot containing the proper fingerprint and vote vector exists in the given precinct for each item selected. Two possible sampling methods exist, which we adapt from [6, 3, 2]. The first option is to sample a fixed number of items from this list. We call this the fixed sample size method. Given N total reported ballots, a minimum of $B = margin/2$ incorrect electronic ballots, and a desired confidence level of c, we require a minimum sample size, n, of:

$$n = \min\left\{ u \mid 1 - \prod_{k=0}^{u-1} \frac{N-B-k}{N-k} \geq c \right\}$$

Alternatively, we may select each electronic ballot independently with probability p, where $p \geq 1-(1-c)^{1/B}$. This latter approach yields a variable sample size and results in marginally more ballots selected on average, but it is more amenable to optimizations, as discussed below. We call this the variable sample size method.

Given n or p, a number of existing papers describe how to securely and efficiently sample from a list of items, and we refer the reader to those papers for greater detail (see [5, 9, 6]).

When a ballot is sampled, auditors feed ballots from that precinct into a scanner. The scanner stops when it observes a match for that ballot's electronically reported fingerprint. All auditors and observers may verify that the vote vectors match, and observers may use their own scanners to verify the fingerprint. Because this check only verifies that a paper ballot exists matching each sampled electronic ballot, observers in this process only need the ability to personally scan the paper ballots that reportedly match the selected electronic ones—not all paper ballots—making the process far more efficient.

Table 1 compares the number of ballots manually reviewed with these methods to the number manually reviewed with precinct-based auditing (using the methods in [16]) for races with margin under 20% in the 2006 Virginia general election. For example, in Virginia's 2006 Webb-Allen U.S. Senate race with a margin of 0.39%, the fixed sample size fingerprinting method requires manual review of 2,337 of 2,370,445 ballots, and the varying sample size method requires review of 2,339 ballots (on average). Precinct-based auditing requires manual review of 1,141,900 ballots (on average).

Additional techniques are possible to reduce the number of ballots to be sampled, potentially resulting in dramatic efficiency gains. For example, we may take the reported contents of ballots into account when determining the probability that we review each of those ballots. Given that all vote vectors must be public for this auditing method, techniques that consider ballot contents are straightforward to apply. See [6] for details.

This auditing process requires that fingerprint-vote vector combinations be released and that scanners be allowed near the ballots following the election. As discussed earlier in this section and in Section 3, these choices may enable certain attack scenarios, so officials must carefully utilize other countermeasures to ensure ballot secrecy. In addition, the practical efficiency and simplicity of this process are unclear and require additional testing. As an alternative, one could check the electronic ballot to paper ballot correspondence using machine-assisted auditing [6], reducing the number of ballots for which the vote combination is revealed publicly (with that technique, only precincts containing sampled ballots must reveal vote vectors). The techniques in the following sections would remain applicable.

4.2 A Paper-to-Electronic Check

Suppose that, in addition to an ability to change electronic records, an adversary has the ability to add paper ballots to the ballot box. This may be true for a number of reasons. For example, a compromised DRE may print additional paper ballots or an official may push extra ballots through the slot of a locked ballot box. The previous check verifies that each electronic ballot has a corresponding paper ballot. If an adversary can add paper ballots, however, the "corresponding paper ballots" might be fraudulently added ballots while the legitimate ballots may be excluded from the electronic results. This would allow an adversary to steal all electronic ballots in a precinct, yet every electronic ballot would have a matching paper ballot (the ballot box would just have many extra unmatching ballots). In this case, we must check not only that each electronic ballot has a corresponding paper ballot but also that each paper ballot has a corresponding electronic ballot. Traditionally, this would require a count of the number of paper ballots in all precincts, but fingerprinting can allow more efficient approaches.

Note that we ignore an adversary that also has the ability to remove paper ballots from the ballot box. Election procedures often dictate that the ballot box should only be unlocked under the watchful eyes of observers. Further, an adversary that can also remove ballots could commit fraud that would be undetectable from the paper and electronic records alone. Such an adversary could arbitrarily modify the paper ballots to match any electronic results.

As in the previous section, we assume use of an optical scan voting machine. When a voter submits her paper ballot, the voting machine records the ballot's fingerprint (but not the vote-vector, so voter privacy is protected). At the end of election day, officials publish both the fingerprints and the voter sign-in sheet. Any interested party can confirm that the reported number of ballots matches the number of voters. As the final step, election officials and other interested parties will have the opportunity to select paper ballots from certain precincts' ballot boxes and ensure that they match the published fingerprints from those precincts. The full version of this paper discusses which precincts to examine and how many ballots to select from those precincts. The key property is that we can achieve the desired level of confidence in the election's outcome as long as any participant selects randomly from the ballot box.

When participants perform this sampling, if they see a ballot not on the reported list of fingerprints for the precinct, this indicates that additional ballots are in the ballot box. The security of this scheme rests on an adversary's inability to produce another ballot with the same fingerprint as a legitimate ballot. These methods rely on difficult-to-duplicate properties of paper and are not possible with serial numbers. If we sample a ballot with serial number 10 from a ballot box, no guarantee exists that another ballot with the same serial number is not in the same box. If we sample a ballot with a given fingerprint from a ballot box, we may reasonably believe that no additional ballots with the same fingerprint are in the same box, even if an adversary has attempted to undermine this property.

4.3 Pre-Scanning Ballots

An additional property that elections officials may wish to check is whether the set of paper ballots in the ballot box is the same as the set of ballots delivered to the precinct on election day. To do so, officials may prescan and publish fingerprints for the paper ballots prior to election day. Immediately prior to the election, election officials, candidates, etc. may verify that these published fingerprints are correct by checking that paper ballots exist matching the published fingerprints and that no extra paper ballots exist (using methods as in the previous two sections). In this way, participants may draw statistically strong conclusions that the set of paper ballots matches the published fingerprints.

When voters submit their paper ballots, the optical scan machine may store fingerprints for the ballots.[7] Following the election, officials may sample fingerprints and paper ballots to verify that legitimate ballots ended up in the ballot box and that the set of reportedly unused ballots were actually unused. Among other things, this check would allow auditors to isolate the set of legitimate ballots in a ballot box, helping to deter attacks that

[7] The machine could even reject ballots with invalid fingerprints, though officials should not trust that the machine performs this check.

rely on slipping in fake ballots.

5 Analysis of an End-to-End Scheme

While our primary focus is on widely deployed voting systems, we briefly discuss the impact of paper fingerprinting on an existing end-to-end voting scheme. A full analysis of these consequences (particularly positive consequences) is somewhat system-specific. Our analysis suggests, however, that end-to-end systems must demonstrate caution if relying on paper.

Scantegrity II. Chaum et al. present a system providing end-to-end verifiability: Scantegrity II [7]. It is an interesting proposal, and it provides for a useful case study.

The basic Scantegrity II system relies on a process similar to traditional optical scan voting, but selection of a candidate causes certain codes written in invisible ink to appear on the paper ballot. Voters may copy the codes for chosen candidates to a paper tab that is to be removed from the ballot and serves as a receipt. These paper tabs contain serial numbers that match serial numbers printed on the paper ballots themselves. When voters cast their ballots, the relationships from voters to codes to candidates is partially destroyed, and the ballots are kept in a locked box (see [7] for additional details and explanation). The system relies on these security measures to prevent an adversary from recovering a voter's ballot or otherwise learning the choices on that ballot.

By keeping paper ballots secure at all times following submission of the ballots, the Scantegrity II system provides a certain level of defense against attacks utilizing paper fingerprinting. If paper ballots are not revealed after voting concludes, an adversary cannot reassociate those ballots with pre-computed fingerprints. Although some risk exists if this assumption of physical security is violated, the serial numbers printed on ballots pose a far greater risk than fingerprints. Overall, the assumptions of the basic Scantegrity II system increase the difficulty of fingerprinting-based attacks, but a violation of these assumptions would raise the possibility of attacks like those against any other optical scan system.

Paper fingerprinting can present a significant benefit in this context, however. One threat to this system is if an adversary can introduce false ballots with the same code corresponding to multiple candidates (see [7]). Using the techniques in Section 4.3, officials may assemble a list of valid ballot fingerprints prior to election day. In the event that a voter receives a fraudulent ballot, the optical scan machine could detect and reject the unexpected ballot.

While fingerprinting can pose serious risks to the Scantegrity II system if assumptions are violated, it also presents an opportunity to improve on existing work. The benefits and drawbacks of fingerprinting will vary dramatically based on details of the end-to-end system. As a general rule, however, any end-to-end system that relies on the uniformity of paper to provide security is at risk. Nevertheless, careful analysis may reveal mitigation strategies and even ways in which paper fingerprinting can strengthen these systems.

6 Conclusion

Paper fingerprinting poses both challenges and opportunities for election officials. This paper outlines several threats to ballot secrecy due to recent advances in paper identification and suggests mitigation strategies to counter these threats. While the most obvious consequences of paper identification are negative, it can also help improve election integrity. Fingerprints can enable an efficient post-election audit process and help detect and prevent additional threats to election integrity.

As technology and algorithms improve, it may be possible for digital cameras and other handheld devices to fingerprint ballots. These new advances will pose additional risks to ballot privacy and should be addressed by future work. For the near future, however, paper will likely remain a critical component of the voting process due to its reliability, cost, familiarity to the public, and ability to stymie many threats to electronic voting systems.

Acknowledgments

We thank Joe Hall, J. Alex Halderman, and Andrew Appel for helpful suggestions and fruitful discussions. We also thank the anonymous reviewers for useful comments.

References

[1] APPEL, A. W. Effective audit policy for voter-verified paper ballots in New Jersey, February 2007. http://www.cs.princeton.edu/~appel/papers/appel-nj-audits.pdf.

[2] ASLAM, J. A., POPA, R. A., AND RIVEST, R. L. On auditing elections when precincts have different sizes. In *Proc. 2008 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '08).*

[3] ASLAM, J. A., POPA, R. A., AND RIVEST, R. L. On estimating the size and confidence of a statistical audit. In *Proc. 2007 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '07).*

[4] BUCHANAN, J. D. R., COWBURN, R. P., JAUSOVEC, A.-V., PETIT, D., SEEM, P., XIONG, G., ATKINSON, D., FENTON, K., ALLWOOD, D. A., AND BRYAN, M. T. Forgery: 'fingerprinting' documents and packaging. *Nature 436* (2005), 475.

[5] CALANDRINO, J. A., HALDERMAN, J. A., AND FELTEN, E. W. In defense of pseudorandom sample selection. In *Proc. 2008 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '08)*.

[6] CALANDRINO, J. A., HALDERMAN, J. A., AND FELTEN, E. W. Machine-assisted election auditing. In *Proc. 2007 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '07)*.

[7] CHAUM, D., CARBACK, R., CLARK, J., ESSEX, A., POPOVENIUC, S., RIVEST, R. L., RYAN, P. Y. A., SHEN, E., AND SHERMAN, A. T. Scantegrity II: end-to-end verifiability for optical scan election systems using invisible ink confirmation codes. In *Proc. 2008 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '08)*.

[8] CLARKSON, W., WEYRICH, T., FINKELSTEIN, A., HENINGER, N., HALDERMAN, J. A., AND FELTEN, E. W. Fingerprinting blank paper using commodity scanners. In *Proc of IEEE Symposium on Security and Privacy* (May 2009).

[9] CORDERO, A., WAGNER, D., AND DILL, D. The role of dice in election audits—extended abstract. In *IAVoSS Workshop on Trustworthy Elections 2006*.

[10] DODIS, Y., OSTROVSKY, R., REYZIN, L., AND SMITH, A. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing 38*, 1 (2008), 97–137.

[11] JOHNSON, K. C. Election certification by statistical audit of voter-verified paper ballots, October 2004. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=640943.

[12] KELLER, A. M., AND MERTZ, D. Privacy issues in an electronic voting machine. In *In Proceedings of the ACM Workshop on Privacy in the Electronic Society (WPES* (2004), ACM Press, pp. 33–34.

[13] METOIS, E., YARIN, P., SALZMAN, N., AND SMITH, J. R. FiberFingerprint identification. In *Proc. 3rd Workshop on Automatic Identification* (2002), pp. 147–154.

[14] NEFF, C. A. Election confidence: A comparison of methodologies and their relative effectiveness at achieving it, December 2003. http://www.votehere.net/papers/ElectionConfidence.pdf.

[15] SALTMAN, R. G. Effective use of computing technology in vote-tallying. Tech. Rep. NBSIR 75-687, National Bureau of Standards, March 1975.

[16] STANISLEVIC, H. Random auditing of e-voting systems: How much is enough?, August 2006. http://www.votetrustusa.org/pdfs/VTTF/EVEPAuditing.pdf.

[17] XIA, Z., SCHNEIDER, S. A., HEATHER, J., AND TRAORÉ, J. Analysis, improvement and simplification of Prêt à Voter with Paillier Encryption. In *Proc. 2008 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '08)*.

[18] ZHU, B., WU, J., AND KANKANHALLI, M. S. Print signatures for document authentication. In *Proc. 10th ACM Conference on Computer and Communications Security* (2003), pp. 145–154.