Privacy Issues for a Voting System with a Modular Voting Architecture

Arthur M. Keller[1], David Mertz[2], and Arnold Urken[3]
[1]UC Santa Cruz and Open Voting Consortium, ark@soe.ucsc.edu; [2]Gnosis Software, Inc., mertz@gnosis.cx; [3]Stevens Institute of Technology, aurken@stevens.edu

Abstract: The Open Voting Consortium has a developed a prototype voting system with a modular voting architecture that includes an open source, PC-based voting machine that prints an accessible, voter-verified paper ballot along with an electronic audit trail. This system was designed for reliability, security, privacy, accessibility and auditability. This paper describes some of the privacy considerations for the system.

1. INTRODUCTION – WHY A SECRET BALLOT?

The requirements for secrecy in elections depend upon the values and goals of the political culture where voting takes place. Gradations of partial and complete privacy can be found in different cultural settings. For instance, in some cantons in Switzerland, voters traditionally communicate their choices orally in front of a panel of election officials.[1] In contrast, in most modern polities, the ideal of complete privacy is institutionalized by relying on anonymous balloting.[2]

The use of secret balloting in elections—where a ballot's contents are disconnected from the identity of the voter—can be traced back to the earliest use of ballots themselves. The public policy rationales for instituting anonymous balloting are typically to minimize bribery and intimidation of the voter. For example, in Athens, Greece during the sixth century B.C.E., Athenians voted by raising their hands "except on the question of exiling someone considered dangerous to the state, in which case a secret vote was taken on clay ballots."[3] In this case, presumably it was deemed necessary to vote via secret ballot to avoid bodily harm to the voter.

Secret ballots, although not always required, have been in use in America since colonial times.[4] The Australian ballot,[5] designed to be uniform in appearance because it is printed and distributed by the government, was adopted throughout most of the U.S. in the late 1800's. Today, approximately one hundred years after most states in the U.S. passed legal provisions for anonymous balloting, a strong sense of voter privacy has emerged as a third rationale. All fifty states have provisions in their constitutions for either election by "secret ballot" or elections in which "secrecy shall be preserved," which has been interpreted by the courts as an implied requirement for secret balloting.[6] West Virginia does not *require* a secret ballot and leaves that to the discretion of the voter.[7] Fourteen states'[8] constitutions do not list "secret" balloting or "secrecy" of elections and/or ballots explicitly. These states have either state laws (election code) or case law (decided legal cases in that state) that mandate secret balloting or interpret the phrase "election shall be by ballot" to mean a "secret ballot."

These cultural values and practices contribute to the sets of user requirements that define the expectations of voters in computer-mediated elections[9] and determine alternative sets of specifications that can be considered in developing open source software systems for elections. The Open Voting Consortium (OVC)[10] has developed a model election system that aims as one of its goals to meet these requirements. This paper describes considerations for ballot privacy in OVC model. We have not developed a formal model of ballot privacy, and the considerations are not necessarily complete.

The OVC has developed its model for an electronic voting system largely in response to reliability, usability, security, trustworthiness, and accessibility concerns about other voting systems. Privacy was kept in mind throughout the process of designing this system. Section 2 of this paper discusses the requirements for a secret ballot in more detail. Section 3 considers how secrecy could be compromised in some systems. Section 4 describes the architecture of the polling place components of the OVC system. Section 5 describes how the OVC handles privacy concerns. While this paper focuses mostly on privacy issues for U.S.-based elections, and how they are addressed in the OVC system, many of the issues raised are relevant elsewhere as well.

2. SECRET BALLOT REQUIREMENTS

The public policy goals of secret balloting[11]— to protect the privacy of the elector and minimize

undue intimidation and influence — are supported by federal election laws and regulations. The Help America Vote Act of 2002[12] codifies this policy as "anonymity" and "independence" of all voters, and "privacy" and "confidentiality" of ballots. It requires that the Federal Election Commission create standards that "[preserve] the privacy of the voter and the confidentiality of the ballot."[13]

The Federal Election Commission has issued a set of Voting System Standards (VSS)[14] that serve as a model of functional requirements that elections systems must meet before they can be certified for use in an election. The VSS state explicitly:

> To facilitate casting a ballot, all systems shall: […] Protect the secrecy of the vote such that the system cannot reveal any information about how a particular voter voted, except as otherwise required by individual State law;[15]

and:

> All systems shall provide voting booths [that shall] provide privacy for the voter, and be designed in such a way as to prevent observation of the ballot by any person other than the voter;[16]

as well as a lengthy list of specific requirements that Direct Recording Electronic voting systems must meet.[17] The basic, high-level requirement not to expose any information about how an individual voted is required of all voting systems before certification and is the most important. The second requirement listed above is a corollary.

It is not sufficient for electronic voting systems merely to anonymize the voting process from the perspective of the voting machine. Every time a ballot is cast, the voting system adds an entry to one or more software or firmware logs that consists of a timestamp and an indication that a ballot was cast. If the timestamp log is combined with the contents of the ballot, this information becomes much more sensitive. For example, it can be combined with information about the order in which voters voted to compromise the confidentiality of the ballot. Such information can be collected at the polling place using overt or covert surveillance equipment—such as cell phone cameras or security cameras common at public schools. As described below, system information collected by the voting system should be kept separated from the content of cast ballots and used in conjunction only by authorized, informed election officials.

3. HOW SECRECY COULD BE COMPROMISED

3.1 A voter's secret identity

When a voter enters a polling place, she enters with a valuable secret: her identity. A secret ballot is not really "secret" in a general sense — it is possible, and even required, for certain recipients to disclose ballots. A secret ballot is "secret" only in the sense that it is blind as to the identity of the voter who cast it. The anonymity of ballots must apply even to most statistical properties of the voters who cast them; a notable exception, however, is in the disclosure of the geographic distribution of voters who vote certain ways in the aggregate. We all know there are "Republican precincts" and "Democratic precincts," and anyone can easily and legally find out which are which.

Complicating matters is the fact that a voter's secret, her identity, *must* be disclosed at a certain stage in the voting process. To be allowed to vote at all, a voter must authenticate her right to vote using her identity, if only by a declaration of purported identity to elections workers. Depending on jurisdiction, different standards of identity authentication apply—some require identification cards and/or revelation of personal information outside the public domain—but in all cases, identity acts as a kind of key for entry to voting. However, legally this key must be removed from all subsequent communication steps in the voting process.

The act of voting, and the acts of aggregating those votes at subsequently higher levels (called "canvassing" in voting parlance) can be thought of as involving a series of information channels. At a first step, a voter is given a token to allow her vote to pass through later stages; depending on the system model, this token may be a pre-printed ballot form, a PIN-style code, a temporary ballot-type marker, an electronic smart card, or at a minimum simply permission to proceed. Although the OVC has not yet settled on a particular token, we will focus on smart cards in this paper, because they have the most serious implications for

privacy. Outside the US, tokens such as hand stamps in indelible ink are also used, particularly to preclude duplicate votes being cast.

Once at a voting station, a voter must perform some voting actions using either pen-and-paper, a mechanical device like a lever machine or a punch card guide, or an electronic interface, such as a touchscreen or headphones-with-keypad. After performing the required voting actions, some sort of record of the voter's selections is created, either on paper, in the state of gears, on electronic/magnetic storage media, or using some combination of those. That record of selections becomes the "cast ballot." Under the Open Voting Consortium system, the paper ballot produced at a voting station undergoes final voter inspection before being *cast* into a physical ballot box.

After votes are cast, they are canvassed at several levels: first by precinct; then by county, district, or city; then perhaps statewide. At each level of canvassing, either the literal initial vote records or some representation or aggregation of them must be transmitted.

3.2 Understanding covert channels

At every stage of information transmission, from voter entry, through vote casting, through canvassing, a voter's identity must remain hidden. It is relatively simple to describe the overt communication channels in terms of the information that actually *should* be transmitted at each stage. But within the actual transmission mechanism it is possible that a *covert* channel also transmits improper identity information.

Covert channels in a voting system can take a number of forms. Some covert channels require the cooperation of collaborators, such as voters themselves or poll workers. Other covert channels can result from (accidental) poor design in the communication channels; while others can be created by malicious code that takes advantage of incomplete channel specification. A final type of covert channel is what we might call a "sideband attack"—that is, there may be methods of transmitting improper information that are not encoded directly in the overt channel, but result indirectly from particular implementations.

For illustration, let us briefly suggest examples of several types of covert channels. One rather straightforward attack on voter ballot anonymity is repeatedly missed by almost every new developer approaching design from a databases-and-log-files background. If the voting channels contain information about the times when particular ballots are cast and/or the sequence of ballots, this information can be correlated with an under-protected record of the sequence of times when voters enter a polling place. We sometimes call this a "covert videotape" attack. In part, this attack uses a sideband: the covert videotaping of voters as they enter; but it also relies on a design flaw in which ballots themselves are timestamped, perhaps as a means to aid debugging.

A pure sideband attack might use Tempest[18] equipment to monitor electro-magnetic emissions of electronic voting stations. In principle, it might be possible for an attacker to sit across the street from a polling place with a van full of electronics, watch each voter enter, then detect each vote she selects on a touchscreen voting station.

Cooperative attacks require the voter or poll worker to do something special to disclose identity. As with other attacks, these covert channels need not rely on electronics and computers. For example, a malicious poll worker might mark a pre-printed blank paper ballot using ultraviolet ink before handing it to a targeted voter. The covert channel is revealed only with an UV lamp, something voters are unlikely to carry to inspect their ballots. A voter herself might cooperate in a covert channel in order to facilitate vote buying or under threat of vote coercion. One such covert channel is to instruct a bought or coerced voter to cast "marked votes" to prove she cast the votes desired by her collaborator. Unique write-in names and unusual patterns in ranked preference or judicial confirmations are ways to "mark" a ballot as belonging to a particular voter.

3.3 Links between registration data and ballots

Since a voter must identify herself when signing in at the polling place, there is the potential for her identity to be tied to her vote. The token given to the voter to allow her to vote may contain her identity. For example, the voter's registration number could be entered into the smart-card writer and then encoded on the smart card that is given to the voter to enable use of a Direct Recording Electronic voting machine. When the voter registration list is given to the polling place on paper, this channel appears less of

an issue. However, if the voter registration list is handled electronically, then the smart card could easily contain the voter's identity. Diebold's stated intent makes this issue a potentially serious privacy risk.

Diebold already has purchased Data Information Management Systems, one of two firms that have a dominant role in managing voter-registration lists in California and other states. "The long-term goal here is to introduce a seamless voting solution, all the way from voter registration to (vote) tabulation," said Tom Swidarski, Diebold senior vice president for strategic development.[19]

4. OVC SYSTEM OVERVIEW

The Open Voting Consortium is developing a PC-based open source voting system based on an accessible voter-verified paper ballot. We mostly describe the components of the system that operate in the polling place.[20] In addition, there are components at the county canvassing site.

4.1 Voter sign-in station

The Voter Sign-In Station is used by the poll worker when the voter signs in and involves giving the voter a "token." It is a requirement that each voter cast only one vote and that the vote cast be of the right precinct and party for the voter. The "token" authorizes the voter to cast a ballot using one of these techniques.

- Pre-printed ballot stock
  - Option for scanning ballot type by Electronic Voting Machine
  - Poll worker activation
- Per-voter PIN (including party/precinct identifier)
- Per-party/precinct token
- Smart cards

The token is then used by the Electronic Voting Machine or an Electronic Voting Machine with a Reading Impaired Interface to ensure that each voter votes only once and only using the correct ballot type.

If the voter spoils a ballot, the ballot is marked spoiled and kept for reconciliation at the Ballot Reconciliation Station, and the voter is given a new token for voting.

4.2 Electronic voting machine

The Electronic Voting Machine (EVM) includes a touch-screen interface for the voter to view the available choices for each contest and select among them. The EVM then prints a paper ballot, which the voter verifies (possibly using the Ballot Verification Station) and places in the ballot box. The EVM is activated by a token, such as a smart card, obtained at the sign-in station. The EVM maintains an electronic ballot image as an audit trail and to reconcile with the paper ballots at the Ballot Reconciliation Station.

4.3 Electronic voting machine with reading impaired interface

The Electronic Voting Machine with Reading Impaired Interface is a PC similar to the Electronic Voting Machine described above which provides auditory output of the ballot choices and selections made and also supports additional modes of making selections suitable for the blind or reading impaired. Whether these features are integrated into a common voting machine with all functionality, or whether there is a separate configuration for the disabled, is an open question. For example, additional modes of input may be useful for those who can read printed materials, but have physical limitations. The idea is to have a universal design that accommodates all voters.

4.4 Ballot verification station

The Ballot Verification Station reads the ballot produced by the Electronic Voting Machine or the Electronic Voting Machine with Reading Impaired Interface and speaks (auditorily) the selections on the voter's ballot. A count is kept of usage, including counts of consecutive usage for the same ballot, but no permanent record is kept of which ballots are verified.

The Ballot Verification Station could also have a screen for displaying the selections. Such an option, enabled by the voter upon her request, would enable a voter who can read to verify that her ballot will be read correctly for automated tallying.

4.5 Ballot reconciliation station

The Ballot Reconciliation Station reads the paper ballots, both cast and spoiled, and reconciles them against the Electronic Ballot Images from the

Electronic Voting Machine or the Electronic Voting Machine with Reading Impaired Interface.

4.6 Paper ballot

The paper ballot is printed by the Electronic Voting Machine or the Electronic Voting Machine with Reading Impaired Interface. It must be "cast" in order to be tallied during canvassing, testing, or a manual recount.

The paper ballot is intended to be easily read by the voter so that the voter may verify that his or her choices have been properly marked. It also contains security markings and a bar code. The bar code encodes the voter's choices, as expressed in the human readable portion of the ballot. The human readable text should be in an OCR-friendly font so it is computer-readable as well. Voters may use the Ballot Verification Station to verify that the bar code accurately reflects their choices. The Ballot Verification Station not only assists sight-impaired and reading-impaired voters in verifying their ballots, but will also give all voters the assurance that the bar-code on the ballot properly mirrors their choices, as represented in the human-readable text on the ballot.

4.7 Privacy folder

The paper ballot contains the voter's choices in two forms: a form that can be read by people and a bar code that expresses those choices in a machine-readable form.

Poll workers may come in contact with the ballot should they be asked to assist a voter or to cast the ballot into the ballot box. In order to protect voter privacy it is desirable to minimize the chance that a voting place worker might observe the voter's ballot choices. A privacy folder is just a standard file folder with an edge trimmed back so that it reveals only the bar code part of a ballot. The voter is expected to take his/her ballot from the printer of the Electronic Voting Machine or the Electronic Voting Machine with Reading Impaired Interface and place it into a privacy folder before leaving the voting booth.

The privacy folder is designed so that the voter may place the ballot, still in its folder, against the scanning station of the Ballot Verification Station to hear the choices on the voter's ballot spoken.

When handed the ballot by the voter, the poll worker casts the ballot by turning the privacy folder so the ballot is face down, and then sliding the paper ballot into the ballot box.

4.8 Ballot box

The ballot box is a physically secure container, into which voters have their paper ballots placed, in order to "cast" their votes. The mechanical aspects of the ballot box will vary among jurisdictions, depending on local laws and customs. Optionally, a perforated tab is removed from the ballot before placing the ballot into the ballot box, and the tab is handed to the voter. The removal of the tab ensures that the ballot cannot be marked "spoiled."

4.9 Box for spoiled ballots

When a voter spoils a ballot, perhaps because the ballot does not accurately reflect her preferences, the ballot is marked spoiled and placed in a box for spoiled ballots for later reconciliation.

5. OVC BALANCES SECURITY, RELIABILITY AND PRIVACY

This section discusses how the Open Voting Consortium is balancing security, reliability and privacy in its electronic voting system.

5.1 Free and open source software

Opening the source code to a voting system — all stages of it, not only the voting station—is a necessary, though not sufficient, condition for ensuring trustworthiness, including the absence of trapdoors and covert channels. For practical purposes, no system that functions as a black box, in which the implementing source code is maintained as a trade secret, can be known to lack covert channels. Any channel with non-optimal utilization includes non-utilized content that is potentially malicious rather than merely accidental — behavior analysis, in principle, cannot distinguish the two.

Of course, free and open source code is not sufficient to prevent covert channels. Sideband channels, in particular, are never exposed by direct examination of source code in isolation; it is necessary to perform additional threat modeling. But even direct encoding of extra information *within* an overt channel can sometimes be masked by subtle programming tricks. More eyes always reduce the risk of tricks hidden in code. Parallel

implementation to open specifications, and message canonicalization also helps restrict channels to overt content.

A frequent criticism of free and open source software is that, while the code is available for inspection, no coordinated inspection is actually conducted.[21] The absence of Non-Disclosure Agreements and restrictive intellectual property agreements makes it possible for a large body of open source developers to inspect the code. Furthermore, in the realm of elections systems, which are mission-critical for a democratic government, open source software could benefit from a specific group of developers who are tasked with recognizing and repairing vulnerabilities. This is a common need in many open source software projects, and in this sense, it might be an appropriate role for a non-profit institution that has delivered such services to other important projects like GNU/Linux, BIND, the Mozilla tool suite and the Apache web server.

5.2 Privacy in the voting token (e.g., smart card)

The token given to the voter to enable her to use the electronic voting machine might contain information that could compromise her anonymity. Indeed, it is not possible to demonstrate the absence of covert channels through black box testing. Thus, analysis of the software is important to show how the data for the smart card is assembled. Above, we considered the benefits of open source software in that numerous people, both inside and outside the process, have the ability to inspect and test the software to reduce the likelihood of covert channels. The hardware that enables smart-card use also includes an interface used by the poll worker (the Voter Sign-In Station). The nature of that interface limits the type of information that can be encoded. Encoding the time of day in the smart card, either intentionally or as a side effect of the process of writing files to the smart card, is a potential avenue for attack. However, the electronic voting machine receiving the smart card knows the time as well, so the smart card is not needed to convey this information.

We propose to encode in the voting token the ballot type and (particularly for multiple precincts at the same polling place) the precinct. The smart card should also be digitally signed by the smart card enabling hardware, so as to help reduce forgeries.

5.3 Printed ballot

The printed ballot contains a human readable version of the voter's selections. After all, that is how it is a voter-verifiable paper ballot. However, the secrecy of the voter's selections is at risk while the voter carries the paper ballot from the electronic voting machine, optionally to the ballot validation station, and on to the poll worker to cast her ballot.

Our approach is to use a privacy folder to contain the ballot. When the voter signs in, she receives the token plus an empty privacy folder. When the EVM prints the ballot, the voter takes the ballot and places it in the privacy folder, so that only the barcode shows. The barcode can be scanned by the Ballot Validation Station without exposing the human readable portion of the ballot. When the privacy folder containing the ballot is given to the poll worker to be cast, the poll worker turns the privacy folder so the ballot is face down and then slides the ballot out of the privacy folder and into the official ballot box. The poll worker thus does not see the text of the ballot, with the possible exception of precinct and (for primaries) party identifiers that may be printed in the margin.

The privacy folder is an ordinary manila folder trimmed along the long edge so that the barcode sticks out.

5.4 Reading impaired interface

The reading impaired interface is used both by voters who cannot read and by voters who cannot see. Having a segregated electronic voting machine used only by the reading and visually impaired can compromise privacy. It is therefore desirable for the electronic voting machines with the reading impaired interface to be used also by those who can read. For example, if all electronic voting machines incorporated the reading impaired interface, then reading impaired voters would not be segregated onto a subset of the voting machines.

It is important that the ballot not record the fact that a particular ballot was produced using the reading impaired interface. Nor should the electronic voting machine record that information for specific ballots. Using a separate voting station for the reading impaired means that the audit trail

is segregated by whether the voter is reading impaired.

Nonetheless, it is useful for the electronic voting machine to maintain some statistics on the use of the reading impaired interface, provided that these statistics cannot identify specific ballots or voters. These statistics could be used to improve the user interface, for example.

5.5 Privacy issues with barcodes

The Open Voting Consortium system design uses a barcode to automate the scanning of paper ballots. Such barcodes raise several possibilities for introducing covert channels.

The prototype/demo system presented by OVC, for example, used a 1-D barcode, specifically Code128. For vote encoding, selections were first converted to a decimal number in a reasonably, but not optimally, efficient manner; specifically, under the encoding particular digit positions have a direct relationship to corresponding vote selections. These digits, in turn, are encoded using the decimal symbology mode of Code128.

Co-author David Mertz identified the problem that even though barcodes are not per-se human readable, identical patterns in barcodes — especially near their start and end positions — could be recognized by observers. This recognition would likely even be unconscious after poll workers saw hundred of exposed barcodes during a day. For example, perhaps after a while, a poll worker would notice that known Bush supporters always have three narrow bars followed by a wide bar at the left of their barcode, while known Kerry supporters have two wide bars and two narrow bars. To prevent an attack based on this kind of human bar code recognition, 1-D barcodes undergo a simple obfuscation of rotating digits by amounts keyed to a repetition of the random ballot-id. This "keying" is not even weak encryption—it resembles a Caesar cipher,[22] but with a known key; it merely makes the same vote look different on different ballots.

In the future, OVC anticipates needing to use 2-D barcodes to accommodate the information space of complex ballots and ancillary anonymity-preserving information such as globally unique ballot-IDs and cryptographic signatures. At this point, we anticipate that patterns in 2-D barcodes will not be vulnerable to visual recognition; if they are, the same kind of obfuscation discussed above is straightforward. But the greatly expanded information space of 2-D barcodes is a vulnerability as well as a benefit. More bit space quite simply provides room to encode more improper information. For example, if a given style of barcode encodes 2000 bits of information, and a particular ballot requires 500 bits to encode, those unused 1500 bits can potentially contain improper information about the voter who cast the ballot.

Just because a barcode has *room* for anonymity-compromising information does not mean that information is actually encoded there, of course. Preventing misuse of an available channel requires complementary steps. Moreover, even a narrow pipe can disclose quite a lot; it only takes about 10 bits to encode a specific address within a precinct using a lookup table. Even a relatively impoverished channel might well have room for a malicious ten bits. For example, if a non-optimal vote encoding is used to represent votes, it is quite possible that multiple bit-patterns will correspond to the same votes. The choice among "equivalent" bit patterns might leak information.

Eliminating barcodes, it should be noted, does not necessarily eliminate covert channels in a paper ballot. It might, however, increase voter confidence as average voters become less *concerned* about covert channels (which is both good and bad). For example, even a barcode-free printed ballot could use steganography[23] to encode information in the micro-spacing between words, or within security watermarks on the page.

5.6 Ballot validation station

The Ballot Validation Station allows reading impaired voters—or anyone—to hear and therefore validate their paper ballots. Since only the barcode of the ballot (and possibly the ballot type—the precinct and party for primaries) is viewable (and as mentioned above, the barcode is obscured), it is best to keep the paper ballot in the privacy folder. So the Ballot Validation Station should be able to read the barcode without removing the paper ballot from the privacy folder. The back of the ballot should have a barcode (possibly preprinted) saying "please turn over," so a Ballot Validation Station will know to tell the blind voter that the ballot is upside down. So that others will not hear the Ballot Validation Station

speak the choices on the ballot, the voter should hear these choices through headphones.

It is useful to know how many times the Ballot Validation Station is used, and how many consecutive times the same ballot is spoken. It is important to assure that ballot-IDs are not persistently stored by the Ballot Validation Station. In particular, to tell how many consecutive times the same ballot was spoken, the Ballot Validation Station must store the previous ballot-ID. However, once another ballot with a different ballot-ID is read, then that new ballot-ID should replace the previous ballot-ID. And the ballot-ID field should be cleared during the end-of-day closeout. The counts of consecutive reads of the same ballot should be a vector of counts, and no other ordering information should be maintained. Inspection of the code together with clear interfaces of persistently maintained records can help assure privacy.

5.7 Languages

Steve Chessin has identified a problem with ballots for non-English speakers. For the voter, the ballot must be printed in her own language. However, for canvassing and manual counts, the ballot and its choices must also be printed in English. However, this approach makes bilingual ballots easy to identify, and that can compromise ballot anonymity if only a small number of voters in a given precinct choose a particular language. Steve Chessin's solution is to have all ballots contain both English and another language, where the other language is randomly chosen for English speakers.[24]

It is important that the Ballot Validation Station handle multiple languages so the voter can choose the language for validating the ballot. To simplify this process, the ballot barcode can include a notation of the second language, but only if that information does not compromise anonymity. Always choosing a second language at random where none is specifically requested reduces the risk. When the ballot's barcode is scanned by the Ballot Validation Station, the voter is given a choice of these two languages for the spoken review of choices listed on the ballot.

5.8 Randomization of ballot-IDs

Under the OVC design, ballots carry ballot-IDs. In our prototype, these IDs are four digit numbers, which provides enough space for ten thousand ballots to be cast at a polling place. We anticipate this ballot-ID length to remain sufficient in production. The main purpose of ballot-IDs is simply to enable auditing of official paper ballots against unofficial electronic ballot images.

The crucial feature of ballot-IDs is that they must not reveal any information about the sequence of votes cast. The prototype and current reference implementation use Python's 'random' module to randomize the order of ballot-IDs. The module uses the well-tested Mersenne Twister algorithm, with a periodicity of $2^{19937}-1$. Seeding the algorithm with a good source of truly random data—such as the first few bytes of /dev/random on modern Linux systems—prevents playback attacks to duplicate ballot-ID sequences.

Because the ballot-IDs are generated at random by each of the electronic voting machines, it is important that two machines do not use the same random ballot-ID. As a result, the first digit (or character) of the ballot-ID in the reference platform will represent the voting machine ID for that polling place.

The remaining 3 digits of the ballot-ID are randomly selected from the range of 000 to 999. A list is maintained of already used ballot-IDs for this electronic voting machine for this election. (One way to obtain such a list is to scan the stored electronic ballot images for the ballot numbers used.) If the random number generated matches an already used ballot-ID, then that number is skipped and a new random number is generated.

5.9 Information hidden in electronic ballot images and their files

The electronic ballot images (EBIs) are stored on the electronic voting machine where the ballot was created. One purpose of maintaining these EBIs is to reconcile them against the paper ballots, to help preclude paper ballot stuffing. The EBIs are in XML format, which can be interpreted when printed in "raw" form.

We prefer not to store the EBIs in a database on the electronic voting machine. A database management system incurs additional complexity, potential for error, and can contain sequence information that can be used to identify voters. On the other hand, flat files in XML format would include the date and time in the file directory, and that is also a potential privacy risk. We can

mitigate this risk by periodically "touching" EBI files electronically during voting station operation, in order to update the date and time of all files to the latest time. The placement order of the files on the disk, however, may still disclose the order of balloting.

Another approach is to store all the EBIs in a single file as if it were an array. Suppose that it is determined that the largest XML-format EBI is 10K bytes. Since there are 1000 possible ballot-IDs for this electronic voting machine, it is possible to create a file with 1000 slots, each of which is 10K in length. When the ballot is to be printed, the random ballot-ID is chosen, and the EBI is placed in that slot in the file, padded to the full 10K in length with spaces (which would be removed during canonicalization). The file can be updated in place, thereby having only the latest date and time. Alternatively, two files can be used, and the electronic voting machine can write to one, wait for completion, and then write to the other. The benefit of this approach is increased reliability of persistent storage of the EBI file. A similar technique can be used to maintain copies of the Postscript versions of the ballots.

When the polling place closes, the electronic voting machine is changed to close out the day's voting. At this time, the EBIs are written as individual flat files in ascending ballot-ID order to a new session of the CD-R that already contains the electronic voting machine software and personalization. Because the EBIs are written all at once, and in order by ascending random ballot-ID, anonymity is preserved.

5.10 Public vote tallying

It is important that the ballots be shuffled before publicly visible scanning occurs using the Ballot Reconciliation System. The ballots will naturally be ordered based on the time they were placed in the ballot box. As described above, the time or sequence of voting is a potential risk for privacy violations.

An illustration of this problem was reported privately to co-author Arthur Keller about a supposedly secret tenure vote at a university. Each professor wrote his or her decision to grant or deny tenure on a piece of paper. The pieces of paper were collected and placed on top of a pile one-by-one in a sequence determined by where each person was sitting. The pile was then turned over and the votes were then read off the ballots in the reverse of that sequence as they were tallied. One observer noted how each of the faculty members voted in this supposedly secret vote.

5.11 Results by precinct

A key approach to ensuring the integrity of county (or other district) canvassing (i.e., vote tallying) is to canvass the votes at the precinct and post the vote totals by contest at the precinct before sending on the data to the county. As a crosscheck, the county should make available the vote totals by contest for each precinct. However, because the county totals include absentee votes, it is difficult to reconcile the posted numbers at the precinct against the county's totals by precinct, unless the county separates out absentee votes (plus hand-done polling place votes). However providing these separations may reduce the aggregation size to impair anonymity. An even worse threat to anonymity arises when provisional ballots are incrementally approved and added to the tally one-by-one.

Posting the vote totals for primary and special elections potentially poses a threat to privacy, as the vote totals may be quite small. However, it is not necessary to post the totals for the non-partisan contests by ballot type (e.g., party). Rather, the tallies for all the non-partisan contests should be posted by precinct without regard to the ballot type. Doing so reduces the number of small totals that compromise ballot privacy.[25]We propose to exclude provisional ballots from the results posted at the precinct. The county tallies by precinct should be separated into a group of votes included in the precinct-posted tally and a group of votes not included in the precinct-posted tally. As long as there is a publicly viewable canvassing of the votes not included in the precinct-posted tally, the issue of voter confidence in the system will be addressed. If that canvassing process involves ballots that have already been separated from the envelope containing the voter's identity, privacy is enhanced.

The totals by precinct are aggregate counts for each candidate. There is no correlation among specific ballots, an important factor to help assure privacy. However, ranked preference voting schemes, such as instant runoff voting, require that the ordering of the candidates must be separately maintained for each ballot. Vote totals are useful

to help assure that each vote was counted, but they do not contain enough information to produce an absolute majority winner. Therefore, vote totals can be posted at the precinct — independent of ranking — and those totals can also be posted at the county. A voter who specifies a write-in candidate for a ranked preference voting race might in principle be doing so as a marker for observation during the canvassing process. To ensure anonymity, write-in candidates whose vote totals are below a certain threshold could be eliminated from the canvassing process. This threshold must be set to avoid distortions of aggregate scores at the county level.

5.12 Privacy in the face of voter collusion

Complex cast ballots, taken as a whole, inevitably contain potential covert channels. We reach a hard limit in the elimination of improper identifying information once voter collusion is considered. In an ideal case, voters cooperate in the protection of their own anonymity; but threats of vote coercion or vote buying can lead voters to collaborate in disclosing—or rather, proving—their own identity. It is, of course, the right of every voter to disclose her own votes to whomever she likes; but such disclosure must not be subject to independent verifications that attack voter anonymity as a whole.

Elections with many contests, with write-ins allowed, or with information-rich ranked preference contests, implicitly contain extra fields in which to encode voter identity. For example, if an election contains eight judicial retention questions, there are at least 6561 possible ways to complete a ballot, assuming Yes, No, and No Preference are all options for each question. Very few precincts will have over 6561 votes cast within them, so a systematic vote buyer could demand that every voter cast a uniquely identifying vote pattern on judicial retentions. That unique pattern, plus the precinct marked on a ballot, in turn, could be correlated with a desired vote for a contested office.

Ballots may not generally be completely separated into records by each individual contest. For recounts or other legal challenges to elections, it is generally necessary to preserve full original ballots, complete with correlated votes. Of course it is physically possible to cut apart the contest regions on a paper ballot, or to perform a similar separation of contests within an EBI. However, doing so is not generally permissible legally.

The best we can do is to control the disclosure of full ballots to mandated authorities, and maintain the chain of custody over the ballots, including the EBIs. A full ballot must be maintained, but only aggregations of votes, per contest, are disclosed to the general public. The number of people who have access to full ballots should be as limited as feasible, and even people with access to some full ballots should not necessarily be granted general access to all full ballots.

5.13 Privacy in electronic voting machines with voter-verifiable paper audit trails

This section discusses other approaches to voter-verifiable paper audit trails. These issues do *not* apply to the design described in this paper — the voter-verifiable paper *ballot*.[26]

Rebecca Mercuri has proposed that Direct Recording Electronic voting machines have a paper audit trail that is maintained under glass, so the voter does not have the opportunity to touch it or change it.[27] Some vendors are proposing that paper from a spool be shown to the voter, and if the ballot is verified, a cutter will release the paper audit trail piece to drop into the box for safekeeping.[28] The challenge with this approach is to make sure that all of the paper audit trail is readable by the voter and does not curl away out of view, and yet that paper audit trails from previous voters are obscured from view. Furthermore, there is the problem that the paper audit trail would fall in a more-or-less chronologically ordered pile. It is also difficult to reconcile the paper audit trail with the electronic ballot images in an automated manner if the paper audit trail cannot be sheet-fed.

Another approach is to keep the paper audit trail on a continuous spool.[29] While this approach has the potential to allow the audit trail to be more easily scanned in an automated fashion for reconciliation, privacy is compromised by maintaining an audit trail of the cast ballots in chronological order. We described above why maintaining order information is a problem for privacy.

Some voters mistakenly confused a paper *trail* with a paper *receipt*, somehow thinking that a

receipt will increase security. These voters do not understand that such a receipt cannot be compared against something to ensure the vote was correctly counted. But a paper *trail* does allow for recounts and manual audits. A further risk of paper *receipts* is the potential for vote selling or coercion.[30]

6. CONCLUSION

We have described the Open Voting Consortium's voting system that includes a PC-based open-source voting machine with a voter-verifiable accessible paper ballot, and discussed the privacy issues inherent in this system. By extension, many of the privacy issues in this paper also apply to other electronic voting machines, such as Direct Recording Electronic voting machines. The discussion illustrates why careful and thorough design is required for voter privacy. Even more work would be required to ensure that such systems are secure and reliable.

ACKNOWLEDGEMENTS

We acknowledge the work of the volunteers of the Open Voting Consortium who contributed to the design and implementation we describe. In particular, Alan Dechert developed much of the design and Doug Jones provided significant insights into voting issues. The demonstration software was largely developed by Jan Kärrman, John-Paul Gignac, Anand Pillai, Eron Lloyd, David Mertz, Laird Popkin, and Fred McLain. Karl Auerbach wrote an FAQ on which the OVC system description is based. Amy Pearl also contributed to the system description. Kurt Hyde and David Jefferson gave valuable feedback. David Dill referred some of the volunteers.

An extended abstract of this paper appeared at the Workshop on Privacy in the Electronic Society on October 28, 2004 in Washington DC, part of ACM CCS 2004 (Conference on Computer and Communications Security). Other papers on this topic are at http://www-db.stanford.edu/pub/keller under electronic voting. More information on the Open Voting Consortium may be found at http://www.openvotingconsortium.org. An earlier version of this paper was published as "Privacy Issues in an Electronic Voting Machine," in *Privacy and Technologies of Identity: A Cross-Disciplinary Conversation*, Katherine J. Strandburg and Daniela Stan Raicu, eds., Springer Science+Business Media: New York, 2006.

REFERENCES

[1] Benjamin Barber, Strong Democracy (Twentieth Anniversary Edition, University of California Press, 2004).

[2] Alvin Rabushka and Kenneth Shepsle, POLITICS IN PLURAL SOCIETIES: A THEORY OF DEMOCRATIC INSTABILITY (1972).

[3] Spencer Albrecht, THE AMERICAN BALLOT (1942) at 9.

[4] In 1682, the Province of Pennsylvania in its Frame of the Government required "THAT all the elections of Members or Representatives of the People, to serve in the Provincial Council and General Assembly … shall be resolved and determined by ballot." (Votes and Proceedings of the House of Representatives of the Province of Pennsylvania. Printed and sold by B. Franklin and D. Hall, at The New Printing Office, near the Market. Philadelphia, Pennsylvania MDCCLII, at xxxi.) In 1782, the legislature of the Colony/State of New Jersey tried to intimidate Tories by requiring viva voce voting. (At that time, about half of New Jersey voted with ballots and the other half viva voce.) They rescinded this in their next session. (Richard P. McCormick, THE HISTORY OF VOTING IN NEW JERSEY 74 (1953). In 1796, the State of New Jersey required federal elections to be by ballot and extended that to state elections the following year. (*Id.* at 106.) In the 1853 pamphlet SECRET SUFFRAGE, Edward L. Pierce recounted Massachusetts' battle to make the secret ballot truly secret. The Massachusetts Constitution in 1820 required elections for representatives to have "written" votes. In 1839, the legislature attacked the secrecy of the written ballot by requiring the ballot to be presented for deposit in the ballot box open and unfolded. In 1851, the legislature passed the "Act for the better security of the Ballot," which provided that the ballots are to be deposited in the ballot box in sealed envelopes of uniform size and appearance furnished by the secretary of the Commonwealth (State of Massachusetts). The battle waged until a provision in the State Constitution made the secret ballot mandatory. (Edward L. Pierce, SECRET SUFFRAGE 7 (1853)(published by the Ballot Society, No. 140 Strand, London, England).

[5] The more general "Australian ballot" is a term used for anonymous balloting using official non-partisan ballots distributed by the government. See Albright 1942 at 26. "The very notion of exercising coercion and improper influence absolutely died out of the country." *See supra* note 3, at 24, quoting Francis S. Dutton of South Australia in J. H. Wigmore's THE AUSTRALIAN BALLOT SYSTEM (2nd ed., Boston, 1889) at 15-23.

[6] For example, The Delaware Supreme Court recognized that the Delaware's constitutional language amounts to an "implied constitutional requirement of a

secret ballot." *Brennan v. Black,* 34 Del. Ch. 380 at 402. (1954).

[7] *See* W. Va. Const. Art. IV, §2

[8] "In all elections by the people, the mode of voting shall be by ballot; but the voter shall be left free to vote by either open, sealed or secret ballot, as he may elect." (W. VA. CONST. ART. IV, § 2 (2003).

[9] Arthur B. Urken, *Voting in A Computer-Networked Environment*, *in* THE INFORMATION WEB: ETHICAL AND SOCIAL IMPLICATIONS OF COMPUTER NETWORKING (Carol Gould, ed., 1989).

[10]The Open Voting Consortium (OVC) is a non-profit organization dedicated to the development, maintenance, and delivery of open voting systems for use in public elections. See http://www.openvotingconsortium.org/.

[11] There are two aspects to anonymous voting. The first is ballot privacy—the ability for someone to vote without having to disclose his or her vote to the public. The second is secrecy—someone should not be able to prove that they voted one way or another. The desire for the latter is rooted in eliminating intimidation while the former is to curb vote buying. The history of these two concepts is beyond the scope of this paper.

[12] The Help America Vote Act of 2002, 42 U.S.C.A. §§ 15301 – 15545 (West, 2004).

[13] *Id.,* § 301(a)(1)(C). (*Also see* §§ 242(a)(2)(B), 245(a)(2)(C), 261(b)(1), 271(b)(1), 281(b)(1), 301(a)(3)(A)).

[14] Federal Election Commission, Voting System Standards, Vols. 1 & 2 (2002), *at* http://sims.berkeley.edu/~jhall/fec_vss_2002_pdf

[15] *Id.* at Vol. 1, §2.4.3.1(b).

[16] *Id.* at Vol. 1, §3.2.4.1.

[17] *Id.* at Vol. 1, §3.2.4.3.2(a)-(e) and §4.5.

[18] *See* http://www.cryptome.org/nsa-tempest.htm (Last visited February 13, 2005)

[19] Ian Hoffman, *With e-voting, Diebold treads where IBM wouldn't*, OAKLAND TRIB., May 30, 2004, *available at* http://www.oaklandtribune.com/Stories/0,1413,82~1865~2182212,00.html

[20] *See* Arthur M. Keller, *et al.*, *A PC-Based Open Source Voting Machine with an Accessible Voter-Verifiable Paper Ballot*, 2005 USENIX ANNUAL TECHNICAL CONFERENCE, FREENIX/OPEN SOURCE TRACK, April 10-15, 2005, pp. 163–174, and *available at* http://www-db.stanford.edu/pub/keller/2004/electronic-voting-machine.pdf

[21] Fred Cohen, *Is Open Source More or Less Secure?* MANAGING NETWORK SECURITY, (July 2002).

[22] *See* http://www.fact-index.com/c/ca/caesar_cipher.html (Last visited February 13, 2005).

[23] Neil F. Johnson and Sushil Jajodia, *Steganography: Seeing the Unseen*, IEEE COMPUTER (February 1998) at 26-34.

[24] It is important to note that the procedure for randomizing the second, non-English language printed on a ballot would have to be quite good. Flaws in the randomization or maliciously planted code could result in the "marking" of certain ballots leading to a compromise of ballot privacy. A simple solution would be to have all ballots printed only in English, and requiring non-English literate voters to use the BVA to verify their vote auditorily. As an alternative for ballots printed only in English, ballot overlays could by provided for each language needed for each ballot type. The overlay could either be in heavy stock paper printed with the contest names with holes for the selections to show through, or it could be a translation sheet showing all the contest names and selections translated into non-English language. In the former case, the ballots would have to be have the layout of each contest fixed, so it would be necessary to have extra spaces when the length of the results vary, such as for pick up to 3 candidates when only 2 were selected. These overlays could be tethered to every voting machine so that voters who read only a specific language could simply place the overlay over their ballot so that she could read their selections as if the ballot was printed in their native language. The overlay approach reduces confusion for English speakers and it also reduces the length of the printed ballot.

[25] See http://libinfo.uark.edu/specialcollections/ACOVH/RoyReed7.pdf, pp. 12–13.

[26] See http://evm2003.sourceforge.net/security.html for the difference between a paper receipt and a paper ballot, and between a paper audit trail and an electronically generated paper ballot.

[27] Rebecca Mercuri, *A Better Ballot Box?*, IEEE SPECTRUM ONLINE (October 2002), *available at* http://spectrum.ieee.org/WEBONLY/publicfeature/oct02/evot.html

[28] For reference, *see* Avanti VOTE-TRAKKER™ EVC308, *available at* http://aitechnology.com/votetrakker2/evc308.html

[29] Press Release, Sequoia Voting Systems, Sequoia Voting Systems Announces Plan to Market Optional Voter Verifiable Paper Record Printers for Touch Screens in 2004, available at http://www.sequoiavote.com/article.php?id=54.

[30] Absentee ballots are incur the risk of vote selling or voter coercion. Another risk of absentee ballots is mass completion, for example at nursing homes.