On the privacy threats of electronic poll books

Stefan Popoveniuc
KT Consulting
Gaithersburg, MD
stefan@popoveniuc.com

John Kelsey
NIST
Gaithersburg, MD
john.kelsey@nist.gov

ABSTRACT

Electronic poll books make the process of verifying that a voter is authorized to vote and issuing her a ballot faster and more convenient. However, they also introduce a privacy risk: if both the electronic poll book and voting machine or optical scanner record the order of sign-ins/votes, or worse, the time at which each voter signs-in and the time at which each ballot is cast, voter privacy can be lost. It is surprisingly difficult to avoid saving such information in some form, for example in event logs on the machines, and still more difficult to verify that no such information is saved. In addition, even the more efficient electronic poll books can act as a bottleneck in the voting process.

We propose a simple technique to address these concerns, by allowing voters to sign-in from home, and print out a bar-coded receipt to be permitted to get their ballot and vote. Using blinded signatures, this barcoded receipt need not leak any information about the voter's identity other than where she is authorized to vote and what ballot she should be given. However, the receipt can contain sufficient information to make it very difficult for a voter to authorize someone else to vote on her behalf, and can support (for example) the ability to respond to challenges by election officials to ensure that only authorized voters are permitted to vote.

1. INTRODUCTION

Protecting the confidentiality of the cast ballots has become increasingly difficult with the introduction of electronic voting equipment. Either by design, or unintendedly, computers keep time stamps of many of the operations they perform, like the creation, modification or access of a file, or a database record. Even the most diligent programmers have a difficult time writing code which is supposed to lose all traces of the order in which some operations happen. This is increasingly difficult because programmers typically reuse a large amount of previously compiled code (like the underlying operating system, the file system, etc), and because, for debugging purposes, programmers have been taught to keep detailed logs for all the operations that their code performs.

If manual, i.e. paper-based, poll books are used, it is easy not to record the order in which the voters checked-in. The check-in judge should only make a check next to the voter's name and should not keep a separate log of the time when the voter arrived. Manual poll books are losing popularity in favor of electronic poll books, which have a series of advantages: no paper is used to print the poll book; the voter can go to any check-in judge (there is no A-M, N-Z division), or even to any precinct if the poll books are networked together; last minute updates to voting roles are easier; and last but not least, there is a trend of doing everything electronically.

The precise order in which the voters arrive at the polling place is likely to be recorded by the electronic poll books at the check-in table. This information can (but should not) be used in conjunction with the order in which ballots are being cast, which may be obtained either from the optical scanner or from the DRE machines. Having the two orders, every cast vote may be easily traceable to a certain voter, or to a small number of voters, violating the secrecy of the ballot. It is considerably more difficult to design and implement a system that does NOT record the time of important actions such as voter check-in or ballot casting, than to have a voting system that DOES record the time stamp. Recording the time stamp usually occurs as a side effect of other actions such as the creation of a file. During development and testing it may be desirable that the system logs all data. Removing this option may be very difficult.

The obvious approach to preventing this linkage is to control the information stored by the electronic poll book and by the voting machine or optical scanner, and to restrict what information may be output by those devices. This is a sensible precaution, but it is hard to do well, and even harder to verify. Even if the specifications state that neither machine shall record the order of sign-ins or votes, and even if the standard reports generated by the machines are always in sort order and contain no time or ordering information, it is all but impossible for any observer to know whether such information has been stored inside the machines, and might be accessed later to reconstruct how everyone voted. When the electronic poll books must be online (for example, to support allowing voters to vote at any convenient polling

place, rather than one specific polling place), this becomes still harder. As with many situations involving privacy, it seems better not to collect the data in the first place than to collect it and try to keep it secret.

There are two ways to prevent the voting system from ever having enough information to link voter identities to ballots: hide the vote, or hide the voter. When hiding the vote, the voting machine does not get to see the clear-text vote that the voter wants to cast, but only an encrypted version of it. Designing such ballots is possible, as proven by systems such as Prêt à Voter [CRS05]. Since only encrypted ballots are available to the voting machine, they cannot provide a tally at the end of the day. Moreover, hand countable paper ballots are not available, and the only way to tally the votes is to decrypt them by a special mechanism (e.g. a mixnet or homomorphic tallying).

This paper focuses on the "hide the voter" approach. The main idea is that voters check-in from home, and they get an anonymous credential that is used as an entry ticket at the polling place. Since the order in which the voters check-in from home is different from the order the voters cast votes at the polling place, and since the identity of the voter is not available to the electronic machines at the check-in table, the correlation between the voter's identity and the ballot she casts is lost. We note that a similar technique might be useful in many other situations where checking in at home makes sense, but where it could represent a privacy violation.

2. PROPOSED TECHNIQUE

Inspired by the airline industry, we suggest to allow voters to check-in from home. Within a given time frame before and during the election, voters are allowed to go to a designated web site, type in their name and address (or username and password, etc), see if their voter registration status is valid, and print a check-in card that has a bar code on it, similar to an airline boarding pass. The card may contain the name of the voter and her address, along with a unique token which is digitally signed.

The check-in card allows the bearer to cast one vote. When the voter gets to the polling place, she presents her check-in card, which is scanned by a bar code reader. The reader checks to see if the digital signature from the bar code is correct, and if the same card has been scanned before. The uniqueness of the token may be used in addition to the personal information that is also contained in the bar code. All the bar code readers can be connected to a central server that helps identify duplicate check-in cards. The check-in judge may ask the voter for a photo ID, check that the information on the check-in card is consistent with the one on the ID, and check that the voter looks like the picture on the ID.

All scanners may be connected to a central server, to be able to prevent double voting. However, the scanners do not need to be connected to a centralized server, since the digital signature on the voter's check-in card can be checked locally, and at the end of the day, all bar code scanners can be connected to a centralized database in order to detect tokens which were reused.

2.1 Blind signatures

A simple way of obtaining an anonymous credential (a token) is by the use of blind signatures [Cha82]. A token consists of a random number that the voter generates, long enough such that it is unique (e.g. a 128 random bit token is unique with very high probability), along with some other information (see section 2.2).

The token is used in a blind signature algorithm that does not allow the voter to derive a second signed token, similar to protocols used in electronic cash[Cha82]. For example, a token t is randomly generated by the voter and a collision resistant one-way function $h(t)$ is blinded by the voter and sent to the authentication server along with the voter authentication credentials (e.g. name and address, or username and password). The server checks the authentication credentials and, if valid, signs the blinded value and marks the voter as being checked-in. Depending on which ballot style the voter is assigned to, the server may use a different private key to sign the blinded token. The server does not have access to the the value, since it is blinded. This offers information-theoretic protection.

The server sends back to the voter the signed, blinded value. The voter un-blinds it, and obtains $h(t)$ which is now signed. She checks that the signed value is $h(t)$ she sent to the server and that the digital signature is correct. The voter prints the signed value $h(t)$ and the token t, and, during voting day, brings them to the check-in station at a polling place.

The signed token is scanned and the voter may be asked to provide some form of identification (e.g. a government-issued photo ID). The check-in judge checks that this is the first time the token has been presented (to prevent the reuse of the token), that the digital signature is valid and that it corresponds to the precinct and ballot style assigned to the voter's address.

The check-in server that signed the tokens has access to the order in which the voters check-in from home, and the electronic poll book at the polling place has access to the order of the signed tokens, but the two machines cannot match the two orders anymore. This is because the order in which the voters checked-in from home is different from the order they come into the polling place, and because the check-in server never got to see the token in clear-text (but only in blinded form). The electronic poll book at the polling place sees the token in clear-text, signed, but never gets to see the identity of the voters (even though the check-in judge does get to check this identity).

The private key that is used by the server to sign the blinded token is unique to the ballot style belonging to the voter. A different private key is used for each ballot style. Since the server has access to the complete identity of the voter, it can easily identify the ballot style corresponding to that voter, and thus use the appropriate private key.

A possible attack against this construction may involve a coercer that collects valid signed tokens from voters and uses them to cast multiple ballots by himself. The same person comes to various polling places multiple times and presents different authorization tokens that are validly signed. This

is possible since the anonymous token is completely independent from the voter's identity, and the check-in judge that verifies the voter's identity and the validity of the token has no way to link the two. The next section presents a specially constructed token that makes this attack impractical.

2.2 A token that is anonymous enough

We present a special construction of the token that includes some attributes of the voter's identity. The token contains partial information about the voter's identity, not enough to uniquely identify a voter, but enough to have her identity validated by a poll worker.

In addition to a unique random number that the voter generates, the token also contains some incomplete information about the voter's identity. For example, the token can contain the first letter of the last name and the last letter of the first name, or it can contain the sex and an interval for the date of birth, or it can say that the last name contains at least 4 vowels and a "T". This way, the poll worker that checks the validity of the token and the ID of the voter, can also check that the identity of the voter is consistent with the partial information in the token. At the same time, the full identity of the voter is still not available to the electronic system that checks the signed token, and the information that is available is not enough to uniquely identify a voter.

To ensure that the blinded token contains partial attributes of the identity of the voter that is doing the check-in from home, a simple zero-knowledge protocol can be used: the voter is asked to create 100 tokens, each with partial information about her identity. The server receives 100 blinded tokens, and, before signing one of them, the server asks the voter to un-blind some random 99 blinded tokens, and checks that all of the opened ones contain partial information about the voter's identity (to which the check-in server has access to). The server can be fairly sure that the un-blinded token which was not opened contains partial information about the same voter. An attacker that presents one identity in clear-text to the server, but includes partial attributes about another identity in the blinded token will be detected with probability 99%. It is easy to see how this probability can be to a value as close to 1 as desired.

Thus, the token contains partial information about the voter's identity, and the poll books at the polling place do get access to this partial information. However, this information should be common to a number of voters, such that the check-in scanner cannot uniquely identify the voter. To be able to sell her voting credential, a voter would have to find another person in her jurisdiction for which this partial information on her token fits with the identity of the fraudulent voter.

2.3 A legal approach

A legal approach may be sufficient to avoid the scenario in which a legitimate voter obtains a valid token via a blind signature protocol and then give the token to another person. It is up to the legal framework to decide how to deal with both the voter who gave her token to somebody else, and with the person who tries to use someone else's token.

To be able to detect an illegitimate use of a token, we can minimally change the protocol by asking the voter to print on a second page the blinding factor used in the blind signature protocol. At random (or if suspicion arises), the check-in judge may ask the voter to provide her blinding factor. The barcode scanner would read the barcode with the blinding factor, contact the sign-in server, and obtain the identity of the voter from the server. The server can obtain this identity, since it has the clear-text token and the blinding factor, and the voter presented her credentials along with the blinded token. Thus the check-in judge can check the identity that was presented to the server for this token, against the identity of the voter who tries to use the token. A mismatch would trigger an alarm and both actors for this fraud may suffer consequences.

Only a small fraction of the voters would have their blinding factor scanned by the check-in judge. For these voters, the time of check-in and their full identity is available to the voting system. The order of the cast ballots does not precisely correspond to the order in which the voters come to the polling place. Rather for one of the voters for which the identity is available to the check-in judge, a number of cast ballots in a certain time frame are possible. Thus ballot confidentiality might still be preserved.

3. MARKETING

Like with many other security products, it may be difficult to convince voters and election officials to adopt our technique solely on the privacy properties which it offers. This section presents additional practical benefits of our proposal, properties which may be used to convince both voters and election officials that it makes their tasks faster and easier. These properties may be presented as the basic features of the new poll book system, and the privacy enhancements would be transparent to the voter and to the election officials.

We identified one of the most common complaints of voters, and we show how our technique addresses it as a side effect.

Waiting in line to vote is one of the most consistent complaints of voters. It is not uncommon for voters to wait in lines for up to 4-5 hours. Voters may be discouraged by the size of the line, and decide not to cast a ballot. Reducing the size of the line by expediting the check-in process is highly desirable.

Assume we have a voting system that uses electronic poll books[1]. The check-in process generally goes as follows. The check-in judge asks the voter for her full name and, using a touch screen and a stylus, types the first three letters of the last name and the first letter of the first name. This usually narrows down the set of registered voters to only a handful of persons. The voter is then asked for the full name, and the check-in judge checks to see if the name is the one that came up on the electronic poll book. The judge also visually checks the sex and the age of the voter. If there is some suspicion that the voter is not who she says she is, the judge can also ask the voter for a photo ID, but typically this does not happen (in other states, showing a photo ID may

[1]For example, the state of Maryland uses electronic poll books

be mandatory). The electronic poll book prints a check-in ticket, which is handed to the voter. The voter has to sign the ticket and take it to the ballot issuing table. The voter surrenders the check-in ticket in exchange for a paper ballot, if optical scan is used, or an activation token, if a DRE is used.

This entire process can be time consuming. The check-in table is often the stop which causes lines to build-up. Other stops are the voting booth where the voter fills in her ballot and eventually the scanner where the voter deposits her paper ballot. In our experience, it is common that the check-in table is the only place where there is a line.

Our technique simplifies the check-in process, and, consequently, the waiting lines at the check-in table would be significantly reduced. The voters that check-in from home may have a dedicated line which would encourage more voters to use this faster option. This would be beneficial for voters, for election officials and for ballot confidentiality too.

3.1 Pre-voting

We briefly mention one other way in which the lines at a polling place can be reduced: pre-voting. This technique is not a contribution of this paper and is independent from the ballot confidentiality problem.

Ron Rivest suggested preliminary voting [Riv05] as a method of shortening the time it take a voter to make her selection on the ballot. From the comfort of her own home, the voter is allowed to fill-in her ballot and print out a representation of her choices. For example, the paper may contain a 2D barcode with her choices. The voter can take all the time she needs when she makes her selections from home. She may consult external sources in order to be better informed about the candidates and issues that she can vote for.

Each voter has to go to a polling place and can bring with her a pre-voted ballot. For simplicity, we assume the voting machine is a DRE and the pre-vote is represented as a bar code. The DRE scans the bar code and pre-fills the electronic ballot with the indicated choices. The voter is allowed to inspect all the choices, and to make any number of changes. The voter may modify all her choices, some of her choices, or none of them before casting her electronic ballot on the DRE.

Even if a coercer forces the voter to bring a certain pre-vote in the booth, the voter is allowed to make any number of modifications and to select her favorite candidates (as opposed to the coercer's favorite candidates).

The ballot that the voter is presented with may contain a large number of races (tens or hundreds) or a large number of candidates in a race. During the time she is at the polling place in front of the voting machine, the voter does not take the time to think about all the choices she can make, but only about a small fraction of them. She may choose to change a couple of choices, but it is expected that the majority of the choices will remain the same, thus reducing the amount of time the voter occupies the voting machine and expediting the voting process.

Rivest underlines that pre-voting is not a precursor for allowing voters to cast a ballot from home, using her own computer and the Internet. If voters do not come in person at the precinct, their ballot would not be cast. This is different from the current Internet voting system implemented in Estonia, where a voter may cast a ballot via the Internet, but can also overwrite it by going to the polling place and casting a subsequent ballot.

4. CONCLUSIONS

We propose a technique which addresses a confidentiality problem caused by the use of electronic poll books in conjunction with any type of electronic voting machines: the identity of the voters is available to the electronic poll books along with the order in which the voters check-in; the options of the voters are available to the voting machines, along with the order in which the ballots are cast. Matching the two orders may result in binding voters' identities with their selections.

To dissociate the two orders, we propose a technique based on blind signatures. We suggest that the anonymous token used in the blind signature protocol contain a small amount of information about the voters identity.

Incidentally, our technique also addresses one of the biggest practical problems for polling places: waiting in lines. Our technique reduces the amount of time a voter spends at a polling place, in particular the amount of time a voter has to wait in line before her credentials are checked and her ballot is issued. Finally, we reiterate Rivest's idea of pre-voting, which may significantly reduce the time a voter spends in front of the voting machine.

5. REFERENCES

[Cha82] David Chaum. Blind signatures for untraceable payments. In *Advances in Cryptology: Proceedings of Crypto'82*, 1982.

[CRS05] David Chaum, Peter Y. A. Ryan, and Steve Schneider. A practical voter-verifiable election scheme. In *In Sabrina De Capitani di Vimercati, Paul F. Syverson, and Dieter Gollmann, editors, ESORICS, volume 3679 of Lecture Notes in Computer Science*, pages 118–139. Springer, 2005.

[Riv05] Ronald L. Rivest. Preliminary Voting - Prevoting. `http://people.csail.mit.edu/rivest/Rivest-PreliminaryVotingPrevoting.pdf`, August 2005.