The "DVSorder" Dominion Ballot Numbering Vulnerability: Recommendations

Ray Lutz, CitizensOversight.org 2022-11-05
raylutz@citizensoversight.org

A privacy vulnerability has been reported in which it may be possible to exploit this vulnerability and link the identity of the voter with their ballot, if the exact order that some or all voters that scanned their ballots is also available. It affects only Dominion Voting Systems (DVS or "Dominion") ImageCast Precinct (ICP) and ImageCast Evolution (ICE) ballot scanners, and only in-person voting. In most cases, it does not affect absentee or vote-by-mail voting. It does not affect in-person voting if the ballots are centrally scanned without resorting them to some known voter order. These details are further described below.

The details of the report are at this site: https://dvsorder.org/

This report also proposes a number of work-arounds that can mitigate or eliminate the vulnerability. We appreciate that this disclosure was made with the best of intentions while carefully weighing the alternative, and so we thank the authors for their work.

However the list of mitigations was incomplete and there was not a balanced risk assessment, thus we believe this announcement raises undo concern. We also recommend that future disclosures omit the details regarding exactly how the Pseudo Random Number Generator (PRNG) operates because providing this information to potential fraudsters is harmful while there is very little benefit to knowing the minute details of the PRNG if you are interested in avoiding the issue.

The following comments apply:

1. **PRIVACY IS IMPORTANT:** Voter privacy is important. There is great benefit to anonymizing ballots, and this fact is undisputed. If voters believe their vote is not private, they may feel pressured to vote with the crowd rather than making an independent vote. Voters may be retroactively persecuted if they voted against a new administration. A number of references regarding this issue are provided below.

2. The Voluntary Voting Systems Guidelines 2.0 (VVSG 2.0)[1] require that the order that ballots are cast cannot be determined:

> **10.2.2-B – No voter record order information**
> The voting system must not contain data or metadata associated with the CVR and ballot image files that can be used to determine the order in which ballots votes [sic] are cast.

The VVSG 1.0[2] has similar requirements:

> **2.3.3.1 Common Requirements**
> To facilitate casting a ballot, all systems shall:
> ...
> b. 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
>
> (and later)
> **3.1.7 Privacy:**
> The voting process shall preclude anyone else from determining the content of a voter's ballot, without the voter's cooperation.

It seems logical that since it is possible to determine the order in which ballots are cast, that these Dominion voting systems are not compliant with the VVSG requirement. However, as these are just "guidelines", it would have to be up to a court to determine whether Dominion is legally responsible. There are general rules in commercial contract law that the products should be "fit for the purpose".

3. **THE METHOD USED BY DOMINION TO ANONYMIZE BALLOTS IS WEAK.**
Dominion used a very weak random number generator which is not sufficient to guard against sophisticated malicious actors.

Yet for opportunistic acts by the general public, it probably will stop most people from linking ballots to voters.

The shuffling done by Dominion can be improved, but to link the ballots to voters, the sequence of voters who use the machine is required. This information is not always easily available. Here, we have a reason to do a better job in the shuffling of the data so

[1] Voluntary Voting System Guidelines VVSG 2.0 -- Requirements for the Voluntary Voting System Guidelines 2.0, February 10, 2021
https://www.eac.gov/sites/default/files/TestingCertification/Voluntary_Voting_System_Guidelines_Version_2_0.pdf
[2] https://www.eac.gov/sites/default/files/eac_assets/1/28/VVSG.1.0_Volume_1.PDF

that even if the order of the voters is available, the data cannot be linked.

4. **UNLIKELY TO BE EXPLOITED IN LARGER VOTING CENTERS.**
In larger voting centers, knowing the order that voters entered the center will likely not be exactly correlated with the order that people will vote. Even with a video of voters exiting, it would be hard to prove that the order of voting was the same as the order of voters exiting.

5. **DVSorder.org Proposed Only Two Mitigation Measures.** The reporters of this problem cited only two mitigation measures:

    a. **Delete the RecordID** in the CSV version of the cast vote record (CVR).

    b. **Use the "dvsanitizer" program**[3] to reassign RecordID numbers to new values which are an encrypted version of the prior entire ballot_id (the term used here to represent the entire ballot identifier, including the TabulatorId, BatchId and RecordId.) This value is encrypted using a seed, which is selected by the user or generated at random, and can be used to decrypt the RecordID and recover the original order. The TabulatorId, and BatchId are not encrypted, which is important for using the records in aggregated reports.

6. **Deleting the record ID may Impact Auditing.**
The deletion of the RecordId is a bad idea, may be illegal, and should be withdrawn as a recommendation. Deleting the RecordId will prohibit any ballot comparison ballot image audit without rescanning the ballots to gain correlated results. Ballot Image Audits (BIAs, for example using https://AuditEngine.org) can provide the highest precision if there is a 1:1 correspondence between the ballot image and the CVR.

    Risk-Limiting Audits (RLAs) using ballot comparison are not possible from these machines anyway, because they need the paper to be correlated with the CVR, and would require a rescan to implement. Other audits that use aggregated results are still possible as long as the ballots can be correctly aggregated to precinct, for example.

    Ballot polling RLAs are still possible but are not as efficient as ballot comparison RLAs. Batch Comparison RLAs are still feasible because the TabulatorId and BatchId are untouched. These together provide a unique batch identifier.

    The RecordId is part of the key field used for indexing the records. Typically, the TabulatorId, BatchId, and RecordId are combined into a single ballot_id, which looks like

        00280_00000_285215

[3] https://github.com/AuburnCyber/dvsanitizer/

Deleting the RecordID in the file (285215 in this case) would require that the entries in the table are reindexed somehow to be able to refer to individual records. Replacing them with numbers like 1, 2, 3, ... is the most obvious. This leaves the records in the same order and may allow the sequence to be reconstructed, so to some extent the problem that we are trying to solve continues to exist.

This option also does not change the filenames on images and it would break the correspondence between the images and the CVR. Breaking this correspondence will not allow ballot-comparison ballot-image audits.

7. **The dvsanitizer program** does a thorough job of changing all the RecordIds to new values and maintains the correspondence with the images because it renames those as well with the same names.[4]

   This is a reasonably good workaround but has some drawbacks.

   a. **LONGER HEX FIELD:** It creates values in hex notation (8 bytes, 32 bits and 16 hex characters + '0x' prefix)

      0xNNNNNNNNNNNNNNNN

      for the RecordId, where the Ns are hex digits (0-9,a-f). This is longer than the normal 6-digit RecordId field, is hexadecimal rather than a decimal integer, is slightly more difficult for humans to use, and some applications may require updates to accept this format. Since the order of the entries is also changed, it would be better to assign the records random numbers in the range from 1 to (the total number of entries), inclusive. This requires more recordkeeping by the code to implement, but is not difficult.

   b. **NOT A LONG-TERM SOLUTION:** The "dvsanitizer" algorithm is not feasible to use as a correction to the Dominion code because it creates non-standard format values.

   c. **DOES NOT CHANGE THE EMS DATA:** The **dvsanitizer** does not change the data in the EMS (Election Management System). Any compromised insider could fully link the ballots to the identity of voters if the exact sequence of voters is also

[4] Ballot image files are supposed to have the same timestamp for the entire election to eliminate the possibility that that metadata could be used to determine the order. We don't happen to know the exact method that Dominion uses to randomize the files on the flash media they use in these voting system scanners. Election Systems & Software (ES&S), the other major vendor of voting systems in the U.S. uses a method where they choose a hash name of the file and sort these into 4096 folders using the first 3 digits of the hex value, and then later, these files are renamed using sequential file names. But inside the flash drive, the files are likely written in order, so it might be possible to discern this if a flash drive was released to the public.

known.

d. **EMS RECORD NUMBERS WILL THEREFORE NOT MATCH PUBLISHED DATA:**
If the data is used for independent image audits (such as by using AuditEngine, see https://AuditEngine.org) and issues were detected in specific ballots, it may be difficult to research these issues in the EMS by staff, if the ballot_id numbers are changed. (As an important side note, AuditEngine does not have any voter list data, so it cannot link voters to ballot data.

e. **CREATES UNUSUAL RecordId VALUES:**
The current code uses AES encryption, combining a random seed with the TabulatorId, BatchId, and RecordId, to create a 16-byte encrypted result, and then it takes the first 8 bytes, resulting in a RecordId that has the format 0xXXXXXXXXXXXXXXXX, i.e. 0x + (16 hex digits). This is distinctive and that means that anyone can easily tell if the records have been corrected.

The original RecordIds are like NNNNNN, a decimal number (but not 0).

This, is not a good idea, for several reasons:

i. There is a small chance of collisions because the encrypted value is truncated to the first half of it. This is a bug and the code should be modified to avoid duplicates in some other way. Indeed the conflict will be rare, but still it is not the best algorithm because of this.
ii. It may be difficult for some programs to deal with the numbers.
iii. It makes it obvious to anyone if the records have been changed.

This last point sounds like it might be a benefit, because then if you forget to keep track, you can look and see whether the files are protected, and then rest easy if they are. But the normal practice with obfuscation of values in files is to provide bogus values that look real. Using values that are in the same format will vastly improve the corrective measure in terms of its power to thwart attack.

Consider a fraudster that wishes to match up voters to their ballots.
If you deploy the version that uses distinctive values that indicate the records have been modified, and if the fraudster finds a CVR that does not have those strange numbers, then they can have some confidence that the numbers provided can be exploited to extract the order, and therefore they might publish their list, reporting "The original file did not have the strange values created by the dvsanitizer app, so we are confident they have not been altered, and therefore we believe this correlation is perfect."

On the other hand, if the values are shuffled so they look like the original values but are not, then a fraudster could only say, "The original file did not have strange values, and we are not sure if they have been modified by the dvsanitizer app, because they are just like the original values," and thus they would not be sure of the correlation. Actually, the shuffled values could be used to create a false correlation. Just the fact that the dvsanitizer app exists, that it shuffles the values and does not make it obvious, will mean that there can never be certainty that the values are the original ones, and thus it thwarts all similar attacks to some degree.

8. **Shuffling RecordId Values is Superior:**

It would be best, of course, if the values looked exactly like the original sequence numbers, but are just shuffled within each batch. FPE (Format Preserving Encryption) is one alternative discussed with the implementers of dvsanitizer, which could be used to create encrypted values that have the same NNNNNN decimal number format, and there is no risk of repeats, but the encrypted values would not have the same characteristics as the defective pseudo-random number generator (PRNG) values. A sophisticated hacker could still determine if they have been altered. If dvsanitizer (or a similar program) provided the same numbers that are just shuffled, then it would not be clear to a fraudster if they have been shuffled, and this protects ALL CVRs, whether they are actually shuffled or not, because the fraudster does not know.

   a. **One way to implement this in code is:**
      i. A random seed is chosen by the operator of the program, using ten-sided dice, for example.

      ii. Scan the CVR (or images) and create a list of RecordIds in a given batch, where a batch is denoted by a unique (TabulatorId, BatchId) pair.

      iii. Use the three values, (an arbitrary seed, TabulatorId, BatchId), to determine a shuffling function to deterministically shuffle the RecordIds, which will not change from their original values, but will be applied to different records. This shuffling function need not be cryptographically perfect, it need only be unpredictable if the seed is not known.

      iv. Apply these to the given batch. Thus the program will need to expand the scope of its operation to a batch at a time rather than just a record at a time, but this is certainly feasible.

      v. The same method is used for images, and it can be applied incrementally, because any incremental releases are generated (at least) a batch at a time.

This shuffling function will produce values that look exactly like the original PRNG values but they are not associated with the same ballots. The values are not changed from those used in the original batch, they are just shuffled and applied to different ballots in the batch. Since it is not possible to tell by looking at the CVR whether it is shuffled or not, it is not possible to prove that they are not, and this protects even CVRs that are not sanitized.

9. **OTHER MITIGATION MEASURES WERE NOT DESCRIBED:**
The DVSorder report does not provide several other mitigation measures that are considered superior, particularly since the recommended mitigations does not shuffle the records in the EMS, and therefore will allow compromised insiders to fully link voters with ballots from these machines, if the voter order is known.

These mitigations are as follows:

a. **RESHUFFLE GROUPS:** Our top recommendation, for situations when these scanners are used only with BMD ballots, is to reshuffle ballots in groups. If the ballots are reordered (shuffled) before they are scanned, this will disrupt the correlation with the order of voters, and will defeat attempts to link the ballots.

Actually, not much shuffling is necessary, even a single cut[5] of the ballots will break the correlation, but additional shuffling can be achieved by flipping half of the cut to reorder all the ballots in that part. This mitigation would be particularly appropriate if the election officials provide camera observation of the voting area.

Because these machines scan BMDs, there is no need for voter feedback during scanning, so the scanning of the ballots can be briefly delayed to allow for shuffling.

A recommended procedure would be as follows:

i. As each voter arrives to cast their ballot. Place it into a privacy-preserving sleeve (like a manila folder), which all look alike.
ii. Put these folders (with ballot inside) in a small box in arbitrary order. In fact, let the voter place it in the box.
iii. When at least a few, perhaps 10 are collected, poll workers should select about half of the folders in arbitrary order, and scan them. Alternatively, as one is added, another random ballot can be selected.
iv. Repeat this until the end of the voting period, and then scan the rest, and close the polls.

Of course, many other shuffling methods may be used, but they are all a variation on this theme. Basically "buffering" up a number of ballots and then using a plan

[5] Interestingly, multiple single cuts of a stack does not improve the shuffle. But unlike cards, ballots can be also "flipped" to reverse the order of all the ballots in that part.

to reorder them. Even if the plan is technically imperfect, it will still add enough uncertainty to the order because it will break the correlation.

Even if there is 24/7 camera observation, it will not be possible to link ballots with voters. The cameras should not be able to follow the ballots through the shuffling operation, of course, and the privacy sleeves must look all the same.

Please note: this recommended mitigation does not apply if you are scanning hand-marked paper ballots which require feedback to the voter for overvotes.

b. **VOTE BY MAIL**
Voters can easily avoid this danger by electing to vote by mail. Normally, VBM ballots arrive in an arbitrary order and are scanned in mixed-precinct batches. It is true that voting at home may not provide absolute secrecy for the voter, (because other people may be present) but if they choose to vote in person using vulnerable machines, the DVSorder vulnerability may be utilized after they have cast the ballot and it is no longer in their control.

Any ballots cast by mail or drop box will be naturally shuffled, and should be further shuffled by election workers, so there is no way to link the voter to their ballot based on the order of voting. Voters can implement this mitigation themselves, and election offices should notify voters of the vulnerability and make VBM voting available to anyone who requests it.

We must admit that there is some risk that compromised insiders may try to glance at ballots from people they know as the ballots are removed from the envelopes. This risk can be limited by procedures and public oversight of the process, which should be easily observable. Some districts have machines that remove ballots from envelopes which may also reduce the risk. It would be difficult for compromised insiders to link voters to their ballots en masse.

c. **CENTRAL SCAN BALLOTS ONLY**
Ballots cast in-person using ballot marking devices (BMDs) can be deposited into secure ballot boxes that do not scan them, and then transported to the central scanning facility, and treat them as if they are validated VBM ballots. There is no feedback needed for the user to correct their marks when they use BMD devices, because the BMD device uses a touch-screen interface and will avoid overvotes. So scanning in the polling place has no real utility for the voter.

Secure ballot boxes are easy to implement. It is done this way in San Diego (which uses Dominion Voting Systems equipment) and Los Angeles (which uses their own equipment called VSAP), and probably many other counties. When ballots are deposited, they fall somewhat randomly into the box, and when returned they can be physically shuffled by staff before scanning. This method

also changes the records in the EMS, and there is no opportunity for compromised insiders to deanonymize ballots. This change may not be feasible for the current elections that have already started, but then the option below can be used.

d. **RESCAN CENTRALLY**
If nonBMD ballots are being scanned by these scanners, or if BMD ballots have already been scanned, paper ballots scanned by the vulnerable equipment can be rescanned in central scanning operations. Any data that may have been collected by those vulnerable machines is therefore disregarded. The paper ballots are the official record and if they are rescanned, it is not necessary to keep the original vulnerable scanned data. This does not require any use of a program to sanitize the data, and completely thwarts malicious actors who are intent on linking ballots to voter identities, including compromised insiders. There is additional delay in both these options because the aggregated totals are not available as fast. In this option, no data is collected from the machines located at remote polling sites.

e. **HYBRID:**
There is a hybrid option, which is to scan the ballots in the polling place, create an unofficial but immediate set of aggregated results, but then rescan all the ballots and create the official CVR and ballot image data using centralized scanners. This unfortunately has several drawbacks, as it still makes the data available to compromised insiders, and it is unclear if the original data from the machines can be effectively disregarded given the strict operational methodology imposed by the voting systems. This is more complicated and would be harder to implement solely to get the first vote totals out quickly, and thus it is not recommended.

f. **RESTRICT VOTER ORDERING DATA:**
It is not a perfect solution, but it does help to limit the usefulness of knowing the order of the ballots if there is no or limited data available regarding the order of casting ballots. This is not required if the mitigations above are used.

More modern poll books may include arrival times (which can help with research on how busy the locations were, for example) and these timestamps may provide some indexing to when the voters voted, but it is imperfect, because the time that is critical is when the ballot was scanned. These days, election officials may be tempted (and the public may demand) to have 24/7 cameras on the voting machines. If used, then this video could be used maliciously to link voters to their data. If there are no timestamps on voter arrival and there are no cameras, then the risk is reduced.

If there are no timestamps and no videos, then the risk is quite low that this

deterministic ordering could be utilized en-masse, and so the mitigation measure of reassigning the numbers has little pay off.

CONCLUSION AND NEXT STEPS

We are circulating this paper to election districts that might be affected by this issue because the disclosure at the website https://dsvorder.org does not fairly represent the risks, does not provide the full set of mitigation options, provides an application that has at least one known bug, and results in ridiculously long RecordId numbers numbers like **0x9ad89e7f9c9656b9d**, which can then be identified as altered, instead of decimal numbers like **983458**. Using these numbers for the RecordId will provide additional information to fraudsters which is not necessary to provide, i.e. that the RecordIds in this file were altered, and thus if the long RecordIds do not exist in the file, they are likely unaltered. It is better to obfuscate using numbers that look legitimate but are shuffled. By doing this, any file, even if the numbers look like they are unaltered, may actually be altered, and this makes any correlation hard to employ with any confidence. An obfuscation program can be run again on the election data and it will shuffle it again. The dvsanitizer program can't be run multiple times. Even if it isn't used, trying to link ballots to voters will have a measure of uncertainty.

We have submitted an issue to the dvsanitizer repository and have had extensive discussions with the implementers. https://github.com/AuburnCyber/dvsanitizer/issues/1

This vulnerability does not provide any ability to impact the accuracy of the count, and thus it is not severe regarding the outcome of the election. For many districts that either use central scanning, are mostly Vote-by-Mail, or do not have 24/7 cameras documenting the order the voters are scanning their ballots, the hazard does not exist or is minimal.

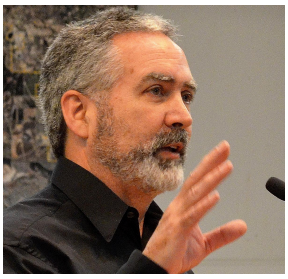
We have provided several important mitigations that can be used instead of the dvsanitizer application, and these are superior because they change the numbers inside the EMS and can thwart attacks by compromised insiders.

We thought this over carefully and came to the conclusion that providing an alternative algorithm that would shuffle the ballots but not be obvious was important, because it makes any CVR useless for outsiders to match voters to their ballots because it is not known if the RecordIds have been shuffled. However, regardless of the algorithm used, compromised insiders might still utilize the ballot RecordIds to deanonymize ballots if the exact order that voters voted is available. The other mitigations, and an eventual fix of this vulnerability by Dominion, will help to completely resolve this issue.

We hope our analysis and solutions will be helpful in this election cycle.

Please also be aware of our Ballot Image Auditing solution. More information at https://auditengine.org

Author: Raymond Lutz



Raymond Lutz is the founder and executive director of Citizens' Oversight Projects, a 501(c)3 nonpartisan nonprofit organization that has been involved in providing oversight to elections for over 15 years. Lutz has a Masters degree in electronics and software engineering, with experience in the document management and printer/scanner/fax/copier industry, and medical device industry. He is the lead developer of AuditEngine.

REFERENCES

The following are a number of interesting references regarding voter privacy.

This 2016 report on ballot privacy is a good resource as we consider the DVSorder hazard.
The Secret Ballot At Risk: Recommendations for Protecting Democracy
https://secretballotatrisk.org/Secret-Ballot-At-Risk.pdf

See also: **EPIC the Electronic Privacy Information Center** -- https://epic.org/
https://epic.org/issues/democracy-free-speech/voter-privacy/

Fact Sheet: Protecting Against Voter Intimidation
https://www.law.georgetown.edu/icap/wp-content/uploads/sites/32/2020/10/Voter-Intimidation-Fact-Sheet.pdf

Preference Falsification -- is the act of misrepresenting a preference under perceived public pressures.
https://en.wikipedia.org/wiki/Preference_falsification

The voting experience and beliefs about ballot secrecy

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209765

ABSTRACT

> New democracies go to great lengths to implement institutional protections of the electoral process. However, in this paper we present evidence that shows that even in the United States—where the secret ballot has been in place for generations—doubts about the secrecy of the voting process are surprisingly prevalent. Many say that their cast ballot can be matched to their name or that others could observe their vote choices while they were voting. We find that people who have not previously voted are particularly likely to harbor doubts about the secrecy of voters' ballots. Those who vote by mail in the privacy of their own homes also feel that others are able to discover their vote choices. Taken together, these

findings suggest an important divergence between public perceptions about and the institutional status of the secret ballot in the United States, a divergence that may affect patterns of voting behavior and political participation.

Excerpts:

1) Specifically, this work finds that those who cast electronic ballots tend to be less confident that their votes are counted properly. There is also evidence that a non-trivial share of the public believes electronic voting technology is susceptible to fraud.

2) As with electronic voting, existing work finds that those who cast mail ballots are less confident that their votes will be counted properly [13] and may have doubts about the effectiveness of precautions taken to ensure the anonymity of a ballot that was sent to them personally at their home.

3) We find no evidence that those who cast ballots early differ systematically from Election Day voters in these top-level perceptions. However, Fig 2 [in the referenced document] shows a clear pattern where those who reported voting by mail harbor greater doubts about secrecy (compared to those who voted in-person on Election Day). They were almost 10 percentage points more likely to say their choices were not kept secret, and approximately 4 percentage points more likely to say it would be not difficult at all for someone to find out about their choices. They were also 15 percentage points more likely to say they thought elected officials could access their voting records to find out who they voted for.