Voter Privacy Flaw Affecting San Francisco's ImageCast Evolution Tabulators

J. Alex Halderman <jhalderm@umich.edu>
Mon 1/9/2023 8:46 PM

To: Jerdonek, Chris (REG) <chris.jerdonek@sfgov.org>;Stone, Robin (REG) <robin.stone@sfgov.org>;Bernholz, Lucy (REG) <lucy.bernholz@sfgov.org>;Crowley, Nancy (REG) <nancy.crowley@sfgov.org>;Dai, Cynthia (REG) <cynthia.dai@sfgov.org>;Livolsi, Renita (REG) <renita.n.livolsi@sfgov.org>

Cc: team@dvsorder.org <team@dvsorder.org>

1 attachments (655 KB)
Letter_to_Jana_Lean.pdf;

This message is from outside the City email system. Do not open links or attachments from untrusted sources.

Dear Members of the San Francisco Elections Commission:

I am Professor of Computer Science and Engineering at the University of Michigan and Director of Michigan's Center for Computer Security and Society. My research focuses on security and privacy, with an emphasis on election systems. I serve as co-chair of the State of Michigan's Election Security Advisory Commission, and I have twice testified to the United States Congress about election cybersecurity issues.

I am writing to draw your attention to a serious voter privacy flaw that affects data published by the San Francisco Department of Elections. Last year, my research group discovered that Dominion ImageCast Evolution (ICE) tabulators, including those in San Francisco, use a flawed method to "shuffle" cast vote records (CVRs) and ballot images. Under some circumstances, the problem would allow members of the public to identify other people's ballots and learn how they voted.

We call this vulnerability "DVSorder." My collaborators and I notified Dominion about the vulnerability in August, and in early October we notified state officials in affected states, including California Chief of Elections Jana Lean via the attached letter. Our letter specifically cited San Francisco as an example of an affected jurisdiction:

> San Francisco has published CVRs and ballot images for all elections since 2019, and data for several other California counties is available elsewhere online. We have confirmed that we can unshuffle ballots in this data that were voted on the affected equipment. . .

Our letter to the state goes on to explain that localities can prevent public exploitation of the flaw by "sanitizing" data they publish to remove or replace the "record ID" field found in CVRs and ballot images.

After notifying all affected states, on October 14 my team published a detailed technical description of the flaw, along with open-source software to help election officials test for and correct the issue, at **https://DVSorder.org**.

A month after we notified California about this problem, San Francisco published data from the November election. I was surprised to find that this data remains vulnerable to DVSorder. The datasets at https://sfelections.sfgov.org/november-8-2022-election-results-detailed-reports do

not appear to incorporate any defense against the privacy flaw, and our publicly available test software indicates that it is possible to unshuffle all of the ballots cast on ICE tabulators and ascertain the date and time each of those ballots was cast. Approximately 128,000 ballots appear to be affected, which
is 8% of the ballots cast in San Francisco during the election.

I attempted to alert the Department of Elections to this issue immediately after the first preliminary dataset was published in November, both via the Department’s online contact form and by Twitter direct messages to @sfvotes. I have yet to receive any response.

San Francisco has earned a reputation as a national leader in election transparency by publishing detailed ballot-level data. It is my hope that the Commission will therefore take the necessary steps to safeguard voters’ privacy with respect to that data by ensuring that the DVSorder issue is addressed in future elections.

I would be happy to make myself available to the Commission and/or the Department to assist with answering any questions.

Sincerely,

J. Alex Halderman
Professor, Computer Science and Engineering
Director, Center for Computer Security and Society
University of Michigan
https://jhalderm.com
734-647-1806

COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

J. Alex Halderman • Professor

October 10, 2022

Jana Lean
Chief of Elections
State of California

Vulnerability Disclosure: Privacy Flaw Affecting Dominion ICP and ICE Tabulators

Dear Chief of Elections Lean:

We are computer scientists at the University of Michigan and Auburn University. We are writing to notify you about a serious privacy vulnerability that affects ballot-level data produced by Dominion ImageCast Precinct (ICP) and ImageCast Evolution (ICE) tabulators. The vulnerability can potentially be discovered and exploited by anyone, without any access to equipment or breach of controls. We believe that California is among 22 states where at least some jurisdictions use these devices.

We were able to discovered the vulnerability, which we call DVSorder, using only publicly available information, so there is an appreciable risk that malicious parties may discover it independently from our work or have already done so. We notified Dominion about the problem on August 23 and informed EAC and CISA on September 2. Since many localities are likely to publish vulnerable data after the November election, **we will be making our findings public this Wednesday, October 12**. Our goal is to inform all entities that operate the ICP or ICE about the risks of publishing vulnerable ballot-level data and about steps they can take to "sanitize" such data so that the vulnerability cannot be exploited by the public. We suggest acting immediately to pause any pending release of vulnerable data until measures are in place to sanitize it.

Many jurisdictions publish ballot-level election data, such as cast-vote records (CVRs) and ballot images. When such data is produced from ICP or ICE tabulators, the DVSorder vulnerability makes it possible for anyone to unshuffle the ballots and learn the order in which they were cast. In some scenarios, knowing the order could make it possible to identify individuals' ballots and determine how they voted. San Francisco has published CVRs and ballot images for all elections since 2019, and data for several other California counties is available elsewhere online[1]. We have confirmed that we can unshuffle ballots in this data that were voted on the affected equipment.

When a ballot is cast on an ICP or ICE, the tabulator assigns it a random-looking "record ID" number. Dominion's EMS software later shuffles the data from the ballots to mask the order in which they were cast[2], but each ballot is still labeled with the original record ID. The vulnerability is that the ICP and ICE are flawed such that they assign ballot record IDs in a predictable manner. This allows anyone to use the record IDs in CVRs or ballot image filenames to determine the order in which the ballots were scanned.

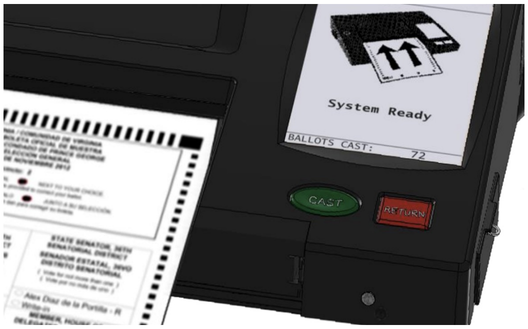
[1] See https://sfelections.sfgov.org/results and https://ordros.com/cvr/California/.

[2] Dominion documentation implies that the shuffled data can be safely distributed without compromising privacy (e.g., https://www.sos.state.co.us/pubs/elections/VotingSystems/DVS-DemocracySuite511/documentation/UG-RTR-UserGuide-5-11-CO.pdf#page=101: exported CVRs entail "[n]o compromise to voter privacy"), as does information Dominion provided during equipment purchasing in some states (e.g., https://www.michigan.gov/-/media/Project/Websites/dtmb/Procurement/Contracts/MiDEAL-Media/008/7700117.pdf?rev=ab164daef7e2459eb741fa7d775f64f0#page=187: "The ballot images are given a random ID number as their file name, and when the images are extracted by the [EMS] application, they are randomized, thus ensuring the ballot images are de-coupled from voter order.").

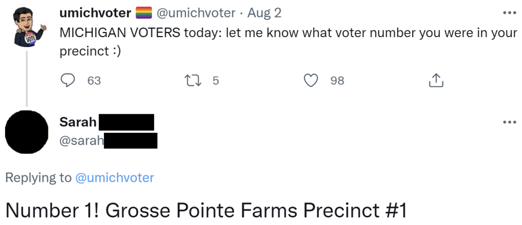
The problem appears to affect all versions of the ICP and ICE, but it does not affect ImageCast Central scanners or ImageCast X DREs. Only data that represents individual ballots and their record IDs is vulnerable; summary results such as statements of votes cast (SoVCs), precinct- or scanner-level totals, election-night result reports, and poll tapes are not susceptible to the privacy flaw.

Here are some scenarios where an attacker could exploit the flaw to identify how individuals voted:

- In most of the country, scanners display a public counter that shows how many ballots have been cast. Anyone can note the counter value when they vote and thereby learn the ballot sequence numbers of people who vote before and after. For example, suppose a man uses the scanner immediately after his neighbor. By noting the counter value, the man can later identify his neighbor's ballot in published CVRs or ballot images and see how she voted.

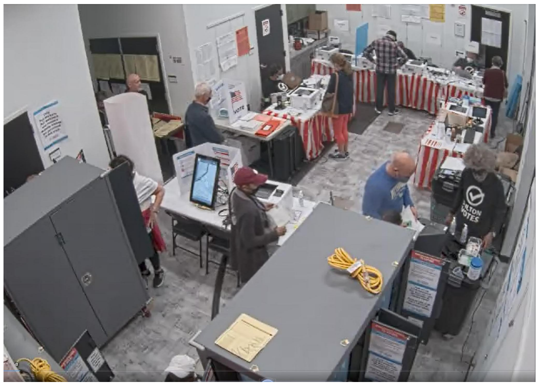


- Poll workers or election observers could similarly note the public counter value to target specific voters. They could also keep a complete record of who uses the scanner, in order, which would allow them to deanonymize all ballots cast at the precinct.
- Some voters publicly disclose their polling places and voter numbers on social media or to others, as in the tweet shown here. As long as the voter has accurately stated their position in the ballot sequence, this would allow anyone to determine how they voted from vulnerable CVRs or ballot images, even for past elections.


umichvoter 🏳️‍🌈 @umichvoter · Aug 2
MICHIGAN VOTERS today: let me know what voter number you were in your precinct :)
63 5 98

Sarah [redacted]
@sarah[redacted]
Replying to @umichvoter
Number 1! Grosse Pointe Farms Precinct #1

- In some states, certain localities publish surveillance footage from inside polling places. (This image is from a day-long video from Georgia.) If the locality also releases vulnerable CVRs or ballot images, anyone can associate each ballot with footage of the voter casting it. Many jurisdictions also treat voter check-in records or poll books as public records. These can heighten the risks posed by the vulnerability, as they often track the order in which voters receive their ballots, which can match or closely approximate the order of casting.



- Some localities, including San Francisco, publish scanner log files (`slog.txt`) from the ICP or ICE. Although these logs by themselves pose little risk to privacy, they can be combined with the DVSorder vulnerability to determine the exact time that each CVR or ballot image was cast (subject to the accuracy of the scanner's internal clock). This provides an additional route to identify voters' ballots.

While this vulnerability is a privacy flaw and does not directly affect the integrity of election results, the secret ballot is itself an important security mechanism. Some voters—particularly the most vulnerable in society—may face real or perceived threats of coercion if the privacy of their votes is not strongly protected. Nevertheless, public access to election data, including CVRs and ballot images, can be a valuable form of transparency that helps uphold voter confidence, so long as the data is carefully prepared to protect privacy.

Fortunately, localities that use the ICP or ICE can prevent the DVSorder flaw from being exploited by the public with no loss of transparency if they take specific steps to "sanitize" ballot-level data before publishing it. Dominion cast-vote records (CVRs) in CSV format use a simple data scheme that can be sanitized manually. To do so, open the `.csv` file in Excel and delete column D, labeled "RecordId", then save the file. Removing ballot record IDs from ballot images (where they they occur in the filenames) and JSON-format CVRs is more labor intensive, so we have developed a software tool to process these files. The tool can sanitize CVRs in `.csv` or `.zip` format and folders of ballots images in `.tif` format. As with any third-party software, jurisdictions should not run our sanitization tool on their EMS computers; instead, we recommend copying vulnerable CVRs or ballot images to an external system and running the tool there. The tool will be available as free open-source software at https://DVSorder.org/ beginning on Wednesday.

Adequately sanitizing ballot-level data makes it just as safe to publish as if the DVSorder vulnerability did not exist. However, even if jurisdictions sanitize the data they make public (or if they do not publish any ballot-level data), the flaw still carries risks. For instance, unsanitized data could be stolen in a data breach or accessed by malicious insiders. Completely mitigating these risks will require Dominion to change the ICP and ICE firmware to use a secure method of generating ballot record IDs. The EAC has told us that Dominion plans to correct the flaw in a future software update, but our understanding is that no patches will be available (at least for federally certified versions) until after the November election. Election officials should contact Dominion for further information and to inquire as to patch availability.

We would be happy to provide whatever information or help we can in support of your efforts to mitigate the vulnerability and strengthen the security and privacy of elections.

Sincerely,

J. Alex Halderman	Drew Springall
Professor Computer Science & Engineering University of Michigan	Assistant Professor Computer Science and Software Engineering Auburn University

Braden Crimmins, Dhanya Narayanan, and Josiah Walker

Student Researchers
University of Michigan