What We Don't Know About the Voatz "Blockchain" Internet Voting System*

David Jefferson
Lawrence Livermore National Laboratory and
Board Member, Verified Voting

Duncan Buell
NCR Professor of Computer Science and Engineering
University of South Carolina

Kevin Skoglund
Chief Technologist
Citizens for Better Elections

Joe Kiniry
Principal Scientist - Galois, Principled CEO and Chief Scientist - Free & Fair

Joshua Greenbaum
Chief Technology Officer
US Vote Foundation

May 1, 2019

Voatz is a recent startup company that is building and operating yet another Internet voting system intended for public elections. The system's major distinguishing features are an elaborate voter authentication system based on automated facial comparison of a photo of a voter's photo ID to a short selfie video, and a back end virtual ballot box in the form a closed, permissioned blockchain.

Some features of the Voatz system have been described in outline, but no detailed technical description has been published in spite of the fact that the system has been used in at least two public elections in West Virginia and may soon be used for another in Denver. Most of the details of the architecture and procedures are apparently confidential, though it is not clear why. The system has not gone through federal certification, or any public certification to our knowledge. The company has not disclosed its source code nor allowed its system to be examined openly by third party experts, as other Internet voting systems have. It has not been subject to open testing in mock elections, again as other Internet voting systems have. The company says it has contracted with several other companies to do a "security audit" of its system, but their reports and findings have not been made public. The company has also been unresponsive to technical questions from outside.

While much of this secrecy might be understandable for an ordinary business product and service, it should not be acceptable in a public voting system whose details should be transparent to voters, candidates, and the public at large. In this document we list a large number of important questions about the Voatz system. We hope the company will be forthcoming and respond to them at some point so the public can more properly evaluate their Internet voting system.

Voter identification and authentication

Voatz has contracted with Jumio, a Palo Alto company in the authentication business, to do its remote voter authentication. The authentication procedure requires the voter, through the

* Author affiliations for identification only.

Voatz smartphone app, to send to Jumio a photo of her driver's license (front and back) or passport photo page, along with a short, live selfie video of the voter's face. Jumio uses machine learning facial comparison software to decide whether the photo of the face on the ID matches the face in the video selfie. If the software decides that it does, the voter is authenticated and her name and address are extracted from the photo ID and returned to Voatz as the true identification of the voter. If the software decides there is no match, there is apparently a backup human comparison of the photo and video, though this has not been explained clearly. Here is some of what we don't know about the authentication procedure.

- Facial comparison is done via machine learning processes that must be trained to match faces. Exactly what training set was used, and how large and diverse was it?
- Driver's license and passport photos are small, and it is only a photo of that photo that is transmitted for facial comparison to the live selfie. Also, the photo on a driver's license or passport can be up to ten years old. Considering these problems for authentication, what are Jumio's rates of false positive and false negative errors in facial comparison?
- Many facial comparison systems have been discovered to have high error rates on minority faces compared to faces from the white majority population. How has Jumio dealt with this issue? How do we know its error rates are the same for all ages, genders, and ethnicities?
- Does a failure in the automated facial comparison always trigger a second facial comparison by a human? If so, what are the false positive and false negative rates for human facial comparison, and how were they measured?
- Jumio offers an authentication service that requires voter eye tracking of a random screen dot to make sure that the video of the voter's face is really live. (A similar system requiring vocal repetition of random words could be used for blind voters.) Why did Voatz not choose that option?
- Between the primary and general elections of 2018 Voatz switched from Yodlee to Jumio for its authentication services. Why was that? Do we expect the relationship with Jumio to continue, or will new authentication contracts be used for different elections or different jurisdictions?

Privacy of sensitive information collected during authentication

Jumio collects an extraordinary amount of sensitive personal information about voters: (1) a live facial video, (2) a photo of a photo ID that contains (3) name, (4) address, (5) birthdate, and (6) driver's license number or passport number, (7) issuance and expiration dates of driver's license or passport, and (8) facsimile of the voter's signature. Furthermore, either Jumio or Voatz also has (9) the voter's smartphone telephone number, (10) a unique ID for the phone device, (11) the IP address(es) used for authentication and voting, and (12) probably the voter's email address.

Jumio's terms of service reads as follows:

User Information License. Customer hereby grants to Jumio a perpetual and irrevocable license to use, reproduce, modify, create derivative works from, distribute, perform, transmit, anonymize and display the User Information (including any rights specifically pertaining to biometric information) necessary to develop, provide and improve the Services, including the right for Jumio to grant equivalent rights to its service providers that perform services that form part of or are otherwise used to perform the Services. Customer further grants to Jumio all necessary rights to use, reproduce, modify, create derivative works from, distribute, perform, transmit and display User Information in an anonymized or aggregated form that does not identify individual persons or organizations (such as, by way of example and not by way of limitation, numbers of verifications) perpetually, in order to compile statistics regarding use of the Services and/or to develop and improve the Services.
Source: https://www.jumio.com/legal-information/terms-and-conditions/v4-3/

This seems to give Jumio the right to retain and use for its own purposes all voter authentication information. This leads to a lot of questions:

- What becomes of all the voter data collected by Jumio? Is it held by Jumio, or Voatz, or election officials, or all three?
- Does the Voatz app use or collect any location data? If so, why?
- What protections prevent this private data from being stolen or sold?
- Does Voatz have a contractual agreement with Jumio guaranteeing that the data will be destroyed, specifically overriding the terms of service quoted above from Jumio's web site? If so, what are the terms of that guarantee?
- What part of this data is destroyed immediately after authentication? What part is destroyed immediately after the election? What part is it held for 22 months? Or longer?
- If the authentication data is stored, where is it stored? Is it stored encrypted? What cryptosystem is used? Who holds the key(s)?
- If the information is destroyed, how do election officials and/or Voatz verify that *all* copies of *all* the information, including backups, are destroyed?
- Exactly how is the authentication data associated with the unique ID later assigned to each voter by Voatz?
- We understand that before Jumio, Voatz had a relationship with Yodlee for authentication services. Was Yodlee used in the West Virginia primary in 2018? If so, what became of the voter authentication data collected by Yodlee? How do we know it was destroyed? Or was it?
- We understand that the Voatz App records every micro-interaction the voter has with it, including timestamped raw data about every touch the voter makes on the screen. What other data is recorded about the voter? Is location information recorded? Since the camera is necessary for authentication, is any photographic or video data recorded? Where is this

data stored? When is it deleted, or is it never deleted? Why is any of this data recorded? What is it used for?

- Is the voter's micro-interaction data transmitted to a server? If so, which server, and who has access to it? Why would that data on voters be collected? Why is there no prominent disclosure of this fact in the Voatz App?

Vote Privacy

It is supposed to be *impossible* to link a voter's identity to the actual vote choices they made so that voters are protected from embarrassment, coercion, and retaliation, and so vote buying, selling, and swapping are not enabled.

- What exactly is the process for separating the voted ballot from any other data that could be used to identify the voter who cast it?
- How do we know that separation is irrevocable — that it is not possible to reconstruct it even approximately?
- As ballots are transmitted to the blockchain, they have to be added to the chain at least in the approximate, if not the exact order they were cast in. How does order preservation not tend to compromise the identity of the voter who cast a ballot if you know the time they cast the ballot?

Phone recognition

After voter authentication, the voter's phone is recognized for future interactions with the Voatz system. How is that accomplished?

- Is the client device sent an authentication token? If so, what information is in the token? Or is an attribute of the client device being stored on the server-side? If so, what attribute?
- How is the token/attribute generated? Is it unique? How likely are collisions? Can it be spoofed?
- If the authentication interaction is with Jumio, how does Voatz recognize the token/attribute as valid and authorized? (Communication? Shared data store? Shared key?)
- Where and how is the token/attribute stored? How is confidentiality and integrity ensured?
- How long does the token/attribute persist? How is it destroyed? Is a new one generated/stored for each election?
- Is the token/attribute the only thing required for future recognition of the phone, or is an additional authentication factor used, e.g. password or SMS?

Voter authorization

Once a voter is authenticated so that the system knows her identity and address, the next step is authorization, i.e. determining that the voter is properly registered to vote in the jurisdiction

she is voting in, determining what ballot she should have based on her address, language preference, and party affiliation (in the case of a primary).

- What kind of access does Voatz have to the voter registration database? Does Voatz have a full copy, or does it forward a separate remote query to the statewide voter registration database for each authorization?
- What prevents Voatz from associating the trove of sensitive personal information about voters collected during authentication with the additional information in the voter registration database about that voter (party affiliation, voting history, and perhaps phone number, SSN, or email address, etc.) to create an extraordinarily valuable and potentially dangerous database usable for identity theft or for illegal political purposes?
- Does Voatz support ballots in multiple languages? If so, which ones? If not, why not?

The Voatz App

The Voatz system is designed so that voters can *only vote from recent IOS or Android smartphones*, and not from desktop or laptop computers. (It is unclear about tablets.) Voting is not done through a web browser.

- Can voters vote from appropriate tablets instead of phones?
- Will Voatz make the source code for the Voatz app available for examination and testing by independent experts without conditions? If not, why not?
- In the Android world voters can download apps from lots of places, not just the Google Play store, so the app binaries are not always vetted by Google, and a lot of malware circulates in the Android ecosystem. What specific technical measures does Voatz take to ascertain that a voter is not voting from a counterfeit version of the Voatz app, one that may have been reverse engineered and modified to behave as malware, possibly changing votes before they are transmitted, or preventing certain voters from voting, or transmitting a copy of the vote choices and the ID of the voter to third parties?
- No two voters are allowed to use the same phone, a rule enforced by associating each voter to the unique ID of the phone used for authentication. Exactly what phone datum is used as the unique ID for this purpose?
- We understand that Voatz also assigns each *voter* a unique ID. Is this true? Is this the same ID as the one used for the phone? If not, exactly how is a unique ID created for a voter? Is it truly random, or a hash of other voter data, or created in some other way? What is the relationship between the phone ID and the voter ID?

Accessibility

One of the common reasons for supporting online voting is the needs of disabled voters.

- Does the Voatz app support an audio interface for blind voters?
- If so, how was it tested?

Ballot transmission

Once the voter has made her choices in the Voatz App the completed ballot is sent over the Internet, eventually arriving at one or more blockchain servers. But so far it has not been documented exactly how this happens.

- Is the ballot transmission done via HTTPS? If not, exactly what protocol and cryptosystem are used?
- Is the ballot directly transmitted from the voter's device to *one of* the blockchain servers, or to *several* of them, or to *all* of them? Or is it transmitted to an intermediate server that then broadcasts to some or all of the blockchain servers?
- What happens if one or more of the targeted servers is down or unreachable at the time of transmission?

Blockchain servers

The server architecture used by the Voatz system has not been fully documented. According to Voatz documentation, 32 blockchain servers were used in the WV general election of 2018, half of them on the Microsoft Azure service and half on Amazon Web Services. Here are some of the things we don't know.

- What security options from Microsoft and Azure have been used to protect the blockchains from penetration attacks or the servers from denial of service attacks?
- What Version of the Hyperledger software does Voatz use for the virtual ballot box? Is it the IBM Fabric framework?
- In West Virginia, was there a separate blockchain for each county, or one unified blockchain for the entire state?
- Exactly what consensus protocol is used to keep the various copies of the blockchain consistent and in agreement? What are its technical properties?
- The 32 servers apparently run exactly the same software configured exactly the same way. There are not multiple implementations with independent flaws. Can we conclude that if there is a bug in one of the Hyperledger instances, the same bug is in all of them?
- Can we conclude that if it is possible to compromise one of the servers, it should be possible to use the same attack to compromise at least the 15 others that are on the same platform, or perhaps all 31 others?
- What exactly is the threat model that makes the use of this kind of blockchain architecture as a virtual ballot box preferable to a replicated database?
- Could an insider at Microsoft or Amazon destroy the whole election by deleting the blockchains and server processes?

Blockchain contents

The contents of the blockchain are not publicly documented, so we don't know what is stored there, and we don't know whether it is encoded in a way that would reveal information about who cast which ballot.

- Exactly what data is stored in the blockchain? Only ballots? Are the ballots keyed to voter IDs or phone IDs?
- How many ballots are stored in each block of the blockchain? A fixed number or a variable number? If it is more than one, how are ballots ordered within a block?
- Is any server log information included in the blockchain, such as time stamps or the IP addresses from which the ballots came? What about authentication and authorization information? Is anything else recorded in the blockchain?

Cryptography

Cryptography is used in several ways in the Voatz system, but nothing seems to be publicly documented.

- Is communication between the voter's device and the blockchain servers done via HTTPS? If not, what encryption scheme is used?
- The ballots are stored encrypted in the blockchain, presumably using AES. Is that correct? If not, what other system is used, and why?
- We have heard that Voatz uses a mixnet of some kind. Is that true? If so, for what purpose? What threats and by what actors does the mixnet protect against?
- It does not seem to make sense to use a mixnet before votes are stored in the blockchain, so presumably it is used for shuffling ballots as they are removed from the blockchain before printing? Is that correct?
- We have heard that Voatz uses an ElGamal cryptosystem. Is that true? If so, what is it used for?
    - For key exchange? If so, is it for communication between phone and Jumio? Between the phone and the blockchain server(s)? Among blockchain servers for the consensus protocol? Some other communication?
    - For digital signatures? If so, for signing what?
    - For homomorphic encryption? If so, is it for tallying, or for a mixnet?
    - For something else?

Transfer to paper

At the close of the election the ballots in the blockchain have to be decrypted, distributed to counties, and transferred to paper ballots for canvass with all the other ballots. How exactly does this happen?

- Do the ballots remain encrypted and get sent to the counties electronically where they are decrypted for transfer to paper? Or are they decrypted, transferred to paper centrally, and then distributed to counties in paper form? Or is there a different process?
- Different counties in West Virginia use different vendors' voting systems. The ballots in the blockchain(s) presumably have to be transferred to paper ballots of several different formats scannable by the counties' vendor-specific systems. What software is used to do that? How was it tested? Or is it done by hand from a screen image?
- Is there any auditing process to make sure the transfer to paper ballots is done correctly? If so, please describe it.

Double voting prevention

At some point there has to be a check to prevent a voter from voting twice, not just through the Voatz system, but via any other means as well. How exactly is double voting prevented?

- When a voter is authorized to vote through the Voatz app, is a check made that she has not already voted? Does that mean that Voatz has access to the county voter registration database (VRDB)?
- Once a voter casts a vote through the Voatz system, is she then recorded in the county VRDB as having voted? Does this mean that Voatz has write-access to county registration databases?
- If a voter "spoils" a Voatz ballot, and does not re-vote, is the indication that she has voted removed from the VRDB?
- How exactly is a voter prevented from voting both by mail-in ballot and also through Voatz, or in person at the precinct and also through Voatz?

"Voter verification" of ballots

In the Voatz system voters are supposed to be able to "verify" that their votes are correctly recorded. This is a misuse of the term "voter verification" because it does not mean what is meant in the literature. What Voatz means is not fully explained, but it seems that a voter can ask through the Voatz app "What votes were recorded in the blockchain for me?", and the vote choices are transmitted back to the voter. There is a lot we don't know about this process.

- Does the request for a copy of a voter's choices to be sent back to her phone go directly to a blockchain server and the answer come directly back, or is there any intermediation? If there is an intermediate process, how does it work?
- How is the ballot being queried identified? Is it keyed on the unique voter ID that was assigned to the voter? Is it keyed to the unique phone ID? Or is it keyed on something else?
- If there is a key that allows finding a particular voter's ballot, how are Voatz or official insiders prevented from looking up a particular voter's ballot?
- How are the voter's choices transmitted back to the voter during verification? By email? Or through the Voatz app?

- If a voter can bring her vote choices back to her phone screen after casting her ballot, how is she protected from coercion or retaliation, and how is the election as a whole protected from vote buying and selling?

Dispute resolution

What is the dispute resolution procedure, i.e. what happens if a voter asserts that her ballot was recorded incorrectly?

- Are disputed ballots marked as such in the blockchain? What is the process for reporting incorrectly recorded ballots, and how are such reports handed? Are statistics kept? What would trigger a forensic investigation into the software?
- In the event a voter believes her vote was recorded incorrectly, is her only recourse to "spoil" the ballot and re-vote?

"Spoiling" a ballot

The Voatz system as of November 2018 allowed voters to "spoil" a ballot after casting it, and to cast a second ballot. This was supposedly without revealing the contents of the first ballot to any official, preserving vote privacy. How exactly does that work?

- Since the first ballot is in the blockchain, it cannot be removed. So presumably a cancellation of some kind for the first ballot is appended to the blockchain, followed by a new ballot. Is that correct?
- How is the ballot to be spoiled found in the blockchain? Is it keyed by the voter's unique ID, or the phone's unique ID? Is it the same mechanism as used for "voter verification"?
- What prevents an insider from using the voter ID or phone ID to impersonate the voter and fraudulently spoil and re-vote on her behalf?
- Can a voter spoil a ballot and re-vote more than once?
- Can a voter spoil a ballot and then *not* re-vote, in effect just withdrawing a cast ballot? And if so, can that voter then vote by mail or at the polls?

Publication of the blockchain

In many discussions of blockchain voting one of the key ideas is that a decrypted version of the blockchain and all of the ballots cast can be made public, so that in principle anyone can verify the integrity of the blockchain and verify the vote counts and the winners. But Voatz has not released the blockchains for either the May 2018 primary election or the November 2018 general election.

- Will Voatz release the blockchains for the 2018 West Virginia primary and general elections? If so, why the delay? If not, why not?

"Auditing" the election

End-to-end auditability is one of the fundamental properties of evidence-based elections. Online elections are inherently unauditable, since there is no trustworthy, indelible record of the intent of the voter as there is with hand-marked paper ballots. Nonetheless, the Voatz literature claims that their system supports "auditability". It is clearly is an abuse of the term, which has a fairly well-defined meaning in the election integrity literature, but Voatz does support an internal consistency check that they call an "audit".

- What exactly is compared to what in the "audit" process?
- Are these "audits" routinely performed for every race in every election? If not, what triggers an "audit"?
- Were the two West Virginia elections in 2018 "audited"? If so, what were the results?
- If an "audit" discovers a discrepancy, and two things that are supposed to be equal are not, what happens?
- How many such discrepancies will trigger a forensic investigation?
- Hypothetically, under what circumstances would an election in which the "audit" was "failed" be considered to be invalid?

"Security audits"

Voatz has stated that it has hired four separate firms to perform "security audits" on their systems, but they have provided little information about these tests.

- What companies did the "security audits"?
- Were these "security audits" done twice in 2018 because of the two separate elections, or only once?
- Exactly what tests and checks were done in these audits?
- What exactly were the results of these audits?
- Will Voatz release the full reports of these audits? If not, why not?

Certification

The Voatz system is not certified to the EAC standards and it cannot be because there are no standards for Internet voting systems, currently or anticipated. Voatz has been floating the idea that it does not need to be certified because it is technically not a voting system, since it does not tabulate votes. But Voatz is certainly *part* of a voting system that *does* tabulate votes. It is a ballot-capturing front end to a complete voting system, and the complete voting system *must* be certified in any state where it is to be used in a statewide or federal election.

- What are Voatz's plans regarding certification?
- Does Voatz acknowledge that it does have to be certified to be used in any public election beyond municipal scale?