A Proof of Concept Study to Increase the Data Integrity of Voting Systems via Redundant Processing

Dr. Bruce D. Holenstein and Victor F. Berutti

Remark Innovations, Inc., Malvern, PA USA

July 10, 2023

Introduction

According to recent polling data, a significant portion of Americans lack faith in the voting process.[1] This distrust, which stems from a perceived or actual lack of integrity in voting systems, has negative effects on the United States.[2] Our group studied the problem and built a Proof of Concept (POC) that makes use of our expertise in data integration and security. We utilized technology from Gravic, Inc., our parent company's, existing product lines. The Gravic Remark Products Group provides optical mark recognition (OMR) software to users around the world, and the Gravic Shadowbase Products Group leads the market in providing data replication technologies and data integration software solutions for mission critical applications.

This whitepaper will discuss the POC as our original contribution to the ongoing discussion about election security. Our proposed solution uses principles of data integrity for mission critical systems to assist in restoring faith in elections.

Our POC goals are as follows:

A. Use paper ballots, hand marked by voters, which are locally tallied at the precinct level,
B. Detect unauthorized ballot creation and copying,
C. Allow for rescanning of ballots with no risk of duplicative counting,
D. Catch hackers who attempt to change results during the counting process,
E. Prevent insider attacks by election workers, and
F. Secure the integrity of both marked and unmarked ballots.

---

[1] Gallup Inc. "Confidence in Election Integrity Hides Deep Partisan Divide." Gallup.com, 2022. https://news.gallup.com/poll/404675/confidence-election-integrity-hides-deep-partisan-divide.aspx.; Ipsos/ABC News. "Ipsos/ABC News Poll (December 27 – December 29, 2021)." Ipsos, 2022. https://www.ipsos.com/sites/default/files/ct/news/documents/2022-01/Topline%20ABC_Ipsos%20Poll%20January%206%202022.pdf.; Murray, Patrick. "National: Public Supports Both Early Voting and Requiring Photo ID to Vote." Monmouth University, 2021. https://www.monmouth.edu/polling-institute/documents/monmouthpoll_us_062121.pdf/.; Pew Research Center. "Two Years After Election Turmoil, GOP Voters Remain Skeptical on Elections, Vote Counts." Pew Research Center, 2022. https://www.pewresearch.org/politics/2022/10/31/views-of-election-administration-and-confidence-in-vote-counts/.; Trafalgar Group. "Nationwide Issues Survey." Trafalgar Group, 2022. https://www.thetrafalgargroup.org/wp-content/uploads/2022/11/COSA-ElectionTrust-Full-Report-1123.pdf.

[2] Organisation for Economic Co-operation and Development. "Trust in Government: Assessing the Evidence, Understanding the Policies." 47th Session of the Public Governance Committee. Paris, France: Organisation for Economic Co-operation and Development, 2013; Alvarez, R. Michael, Jian Cao, and Yimeng Li. "Voting Experiences, Perceptions of Fraud, and Voter Confidence." *Social Science Quarterly* 102, no. 4 (July 2021): 1225–38. https://doi.org/10.1111/ssqu.12940.

Data Integrity in Mission Critical Systems

In 2017, the Department of Homeland Security officially designated voting systems as part of the nation's critical infrastructure.[3] This designation is reserved for systems "so vital to the United States that their incapacitation or destruction would have a debilitating effect on security, national economic security, national public health or safety, or any combination thereof."[4]

This designation indicated that voting systems are mission critical systems. Though voting systems are critical infrastructure and certainly fall into this category, the methods used to process votes is less secure than similar mission critical systems in other industries. There are three basic components of a mission critical system: **Reliability, Availability, and Scalability (RAS).**

> "**Reliability** is a measure of how well a system returns the same correct, consistent, and uncorrupted results each time, and relies on the underlying integrity of the database, application, and system components;
>
> **Availability** is the percent of uptime achieved by the application in servicing users; and
>
> **Scalability** is the capability to add resources when needed to handle the application load, and to return those resources when no longer needed."[5]

Proof of Concept (POC) Overview

The solution chosen for this POC combines trusted optical mark recognition and database replication technologies into a *Validation Architecture*[6] to provide high levels of Reliability, Availability, and Scalability (RAS). Validation Architectures redundantly process data in multiple data centers so that it can be compared and validated in real-time to protect data integrity. It is key technology for mission critical systems.

Our solution includes two main parts:

- Local precinct operations set up to attain a secure voter-facing *front-end*,
- Plus, a high data integrity processing *back-end* to ensure the correct counting of ballots.

Local Precinct Balloting Front-End

We have identified the following workflow for a system that accomplishes a form of "Balloting Gold Standard" that can be replicated in the diverse set of local precincts by those who administer elections:

---

[3] U.S. Department of Homeland Security. "Election Security." Homeland Security, 2023. https://www.dhs.gov/topics/election-security.

[4] Cybersecurity & Infrastructure Security Agency. "Critical Infrastructure Sectors." CISA. Accessed December 31, 2022. https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors.

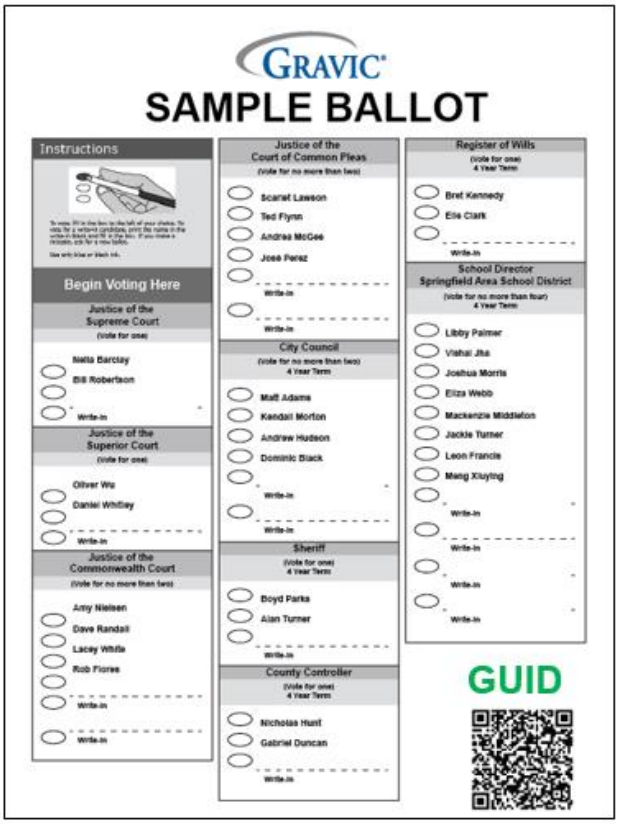
[5] Holenstein, Bruce, Paul J. Holenstein, and Victor Berutti. "New Data Integrity Architectures for Mission Critical Systems." *The Connection*, 2021. https://connect2nonstop.com/new-data-integrity-architectures-for-mission-critical-systems/.

[6] Ibid.

1. Use printed paper ballots on paper that is not easy to copy. The paper may have a watermark or other security features that make forgery difficult.
2. Each ballot is printed at a central location(s) with a "globally unique ID" (GUID) in the form of an alphanumeric string. The string may be printed or placed as a barcode on the ballot. Prior to the election, a Ballot Master Database would be created containing all the valid GUIDs for ballots that are to be used in the election.
3. As illustrated in Figure 1 below, once the polls open, ballots are hand-marked by voters and the candidate selections are readable by a human. The voter then proceeds to the voter-facing scanner and places the ballot in the voter-facing scanner.
4. The scanner scans the marked ballot, and sends an image of it to our Remark software application running on the computer inside the voter-facing scanner.
5. The Remark software processes the image of the marked ballot and validates the ballot.
6. If valid, the vote is counted and a time/date stamped image of the ballot is written to the optical disk or other write-once, un-erasable media for forensic and auditing purposes. However, if a problem such as an over or under vote is detected, the voter will be notified and the ballot returned to the voter for self-adjudication.
7. After the polls close, reports are created for the results from that voter-facing scanner. The report can be printed, or results can be provided to the poll administrator who can pass it to a central counting location. The tallies may be called or faxed into the central office.
8. Optionally, all blank unused ballots may be scanned after the polls close to prevent their use for fraudulent purposes. The GUIDs of the unused ballots will be recorded as "unused."
9. All precinct scanning equipment is housed in tamper-proof locked containers that collect the scanned ballots. The equipment is shielded to reduce or eliminate all electromagnetic and optical emissions. It is also air-gapped so it operates isolated from outside influences. No connections to the internet, phone system, Wi-Fi networks, nor Bluetooth devices are allowed.

Ballot Details

Ballot integrity is key to election security, so voting systems must be able to determine if a ballot has been copied or processed more than once during the election. In this POC, each ballot has a GUID printed on it as illustrated in Picture 1. The term GUID stands for Globally Unique Identifier, an analogous term to UUID (Universal Unique Identifier). A GUID can be a random alphanumeric string printed in non-human readable form as a 1-D or 2-D barcode. The GUIDs could include additional embedded information such as the ballot style or the page number in a case of a multi-page ballot. Ballot GUIDs are generated in the Remark application and stored in the Ballot Master Database when they are printed on a valid ballot.



Picture 1. Sample Ballot with 2-D

Ballot Master Database Details

A Ballot Master Database would need to be created and housed securely in each data center with its contents kept secret. The office creating the ballot GUIDs (we envision that of the Chief Election Officer) would generate GUIDs to support the expected requirements for the election. In the case where the state provides preprinted ballots to the precincts, the GUIDs for all ballots distributed to that precinct should be known before delivery and noted in the Ballot Master Database. In a case where ballots are printed on demand at the precinct, a secure request would be made to return an unused GUID to the printing software and that GUID would be printed on the ballot. The Ballot Master Database would be updated to note that this particular GUID had been used. Once used, that GUID could not to be used again without detection.

Local Precinct Operation Details

When a voter enters the precinct, election administrators give her a paper ballot that includes a unique GUID from the Ballot Master Database. The voter then fills the ballot and deposits it into the voter-facing scanner. The scanner produces an image of the ballot and sends it to a computer securely housed within the voter-facing scanner. If the ballot is error free, the computer would count it and secure it within an internal ballot vault. A time/date stamped image of the ballot would be written to a write-once, un-erasable media for forensic and auditing purposes.

If the computer detects a problem, such as an over or under vote, it will notify the voter and return the ballot for correction. The software would store an image of the rejected ballot with a "rejected" flag but would not include it in any tabulations that follow. A spoiled ballot would be

replaced with a new ballot containing a new GUID. A fixed ballot would be scanned again and processed as described above, as it is then a valid ballot.

Once the polls are closed, any unused ballots could be scanned and marked with an "unused" flag. This step ensures that the ballot master database can account for all ballots and GUIDs produced for the election. The software in each voter-facing scanner can produce results of all valid ballots scanned by that scanner and can raise an alarm should a GUID have been associated with more than one valid scanned ballot.

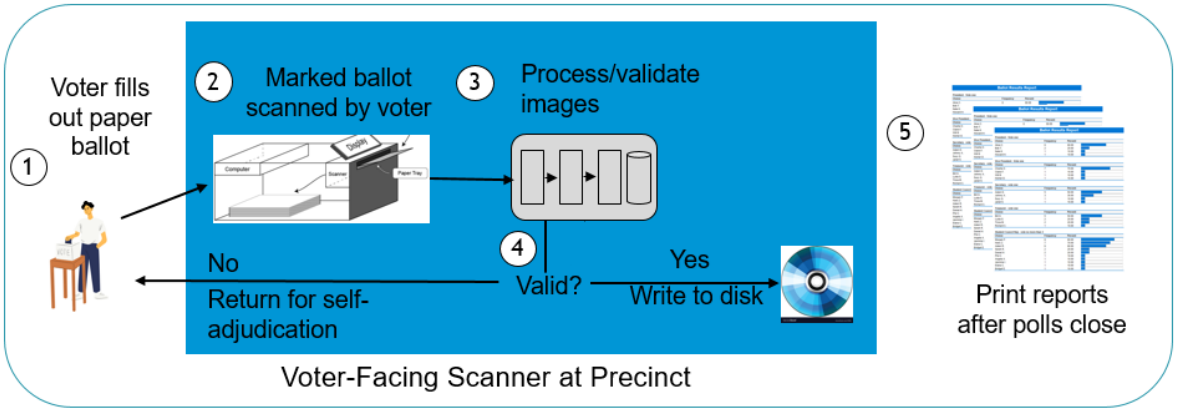


Figure 1 – Voter-Facing Scanner using Remark OMR® Software with Manual Reporting of Results at the Precinct

High Data Integrity Back-End Processing

We have identified the following workflow for a high data integrity back end system which accomplishes a Balloting Gold Standard. The detailed redundant back-end processing steps that ensure complete data integrity will now be described.

Step 1. Batches of physical, secure paper ballots are preprinted before the election with unique IDs (GUIDs), which are recorded in both copies of the Ballot Master Database in systems located at independent processing centers.

Step 2. As previously described, ballots are marked by the voters at a precinct, and are submitted for scanning at the voter-facing scanner where the images are saved to secure storage, probably an optical disk.

Step 3. After the polls are closed and voter-facing scanners are shut down, systems at the precincts are connected to encrypted lines. They then send the images of the scanned ballots to two separate Remark processing instances. At the Processing Centers are separate instances of the ballot counting software, Remark Instance 1 and 2, operating on private secure servers. There is an isolating firewall separating the Processing Centers and no uncontrolled method of communication exists between them. As shown in the POC architecture in Figure 2, precincts on the left side of the figure are geographically distributed as needed and are used to handle the ballot volume.

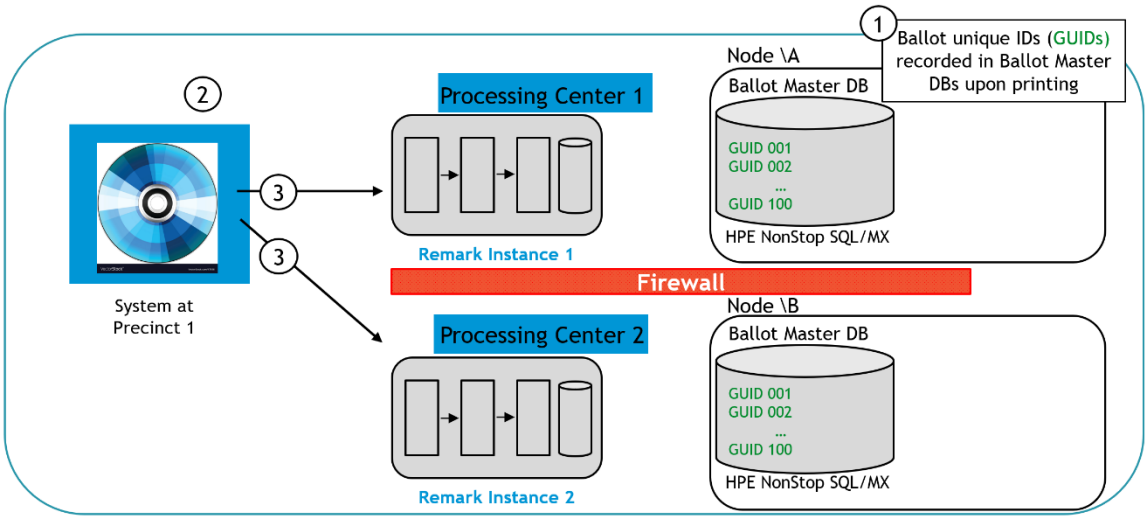


Figure 2 – Validation Architecture: Steps 1-3

Step 4. As shown in Figure 3, each Remark Instance processes the ballot images, independently operating on separate networks.

Step 5. The Remark Instances update their respective Ballot Master Database with the ongoing ballot recognition results.

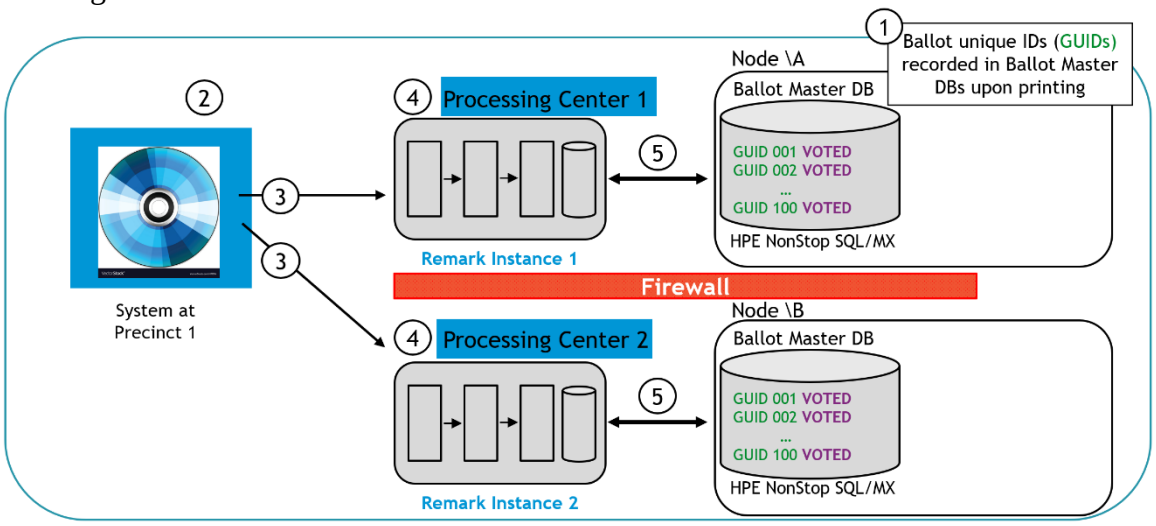


Figure 3 – Validation Architecture: Steps 4-5

Step 6. As shown in Figure 4, all changes to the Ballot Master Databases are recorded in durable Transaction Logs. Validation Architecture components implemented with Shadowbase software on each Ballot Master Database node generate a hash value for each transaction batch of ballots and then exchange the value with each other.

Step 7. If the results match after the comparison, then the results are considered validated in the Ballot Master Database. This step verifies that the Ballot Master Databases are in-sync and that no hardware faults, hackers, insiders, or other threat corrupted the vote tallies.

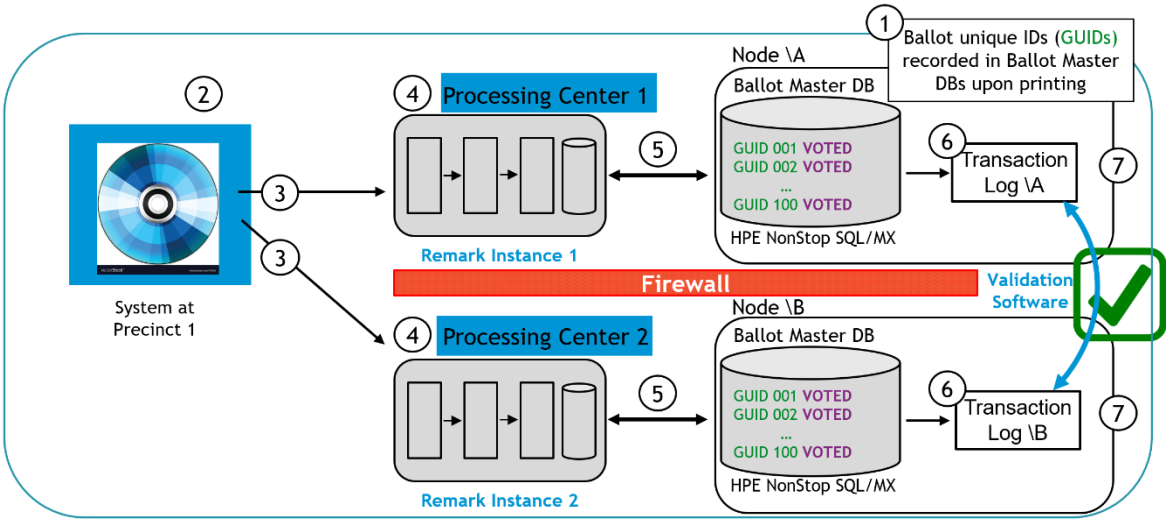


Figure 4 – Validation Architecture: Steps 6-7

However, if the results do not match, then the mismatch is reported and the ballot batch is marked for further review by appropriate authorities as shown in Figure 5.

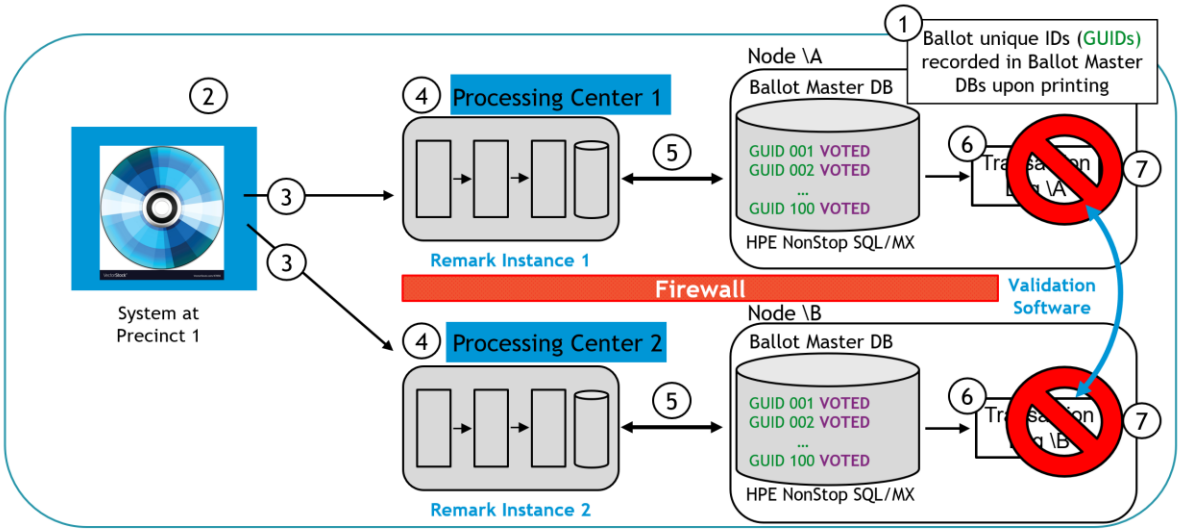


Figure 5 – Validation Architecture: Detecting Ballot Mismatches

Using such a system to accumulate and process ballots not only allows redundant processing but also secures the chain of custody from the local precinct to the state. It eliminates the need to move physical ballots to a centralized counting facility, during which ballots could be misplaced or tampered with. The redundant data centers also ensure that no one can alter results by scanning in extra ballots after polls close because the system counts all votes two or more times and would flag and report discrepancies across counts.

The RAS principles described herein guarantee a series of safeguards against hacking that can detect potential threats at numerous points in the tabulating and auditing processes. Many voters fear that hackers could manipulate elections, but redundant processing means they would have to corrupt two or three different systems simultaneously without raising any alarms. If the system is correctly deploying RAS principles, this would be nearly impossible. Even in that very unlikely scenario, both the paper ballots and ballot images would remain available to election administrators to cross reference and recount during an audit.

Summary

By applying some of the key principals associated with mission critical applications, voting systems can be made more secure and trusted. The solution outlined in this white paper would:

- Catch unauthorized ballot copying and creation of fake ballots,
- Allow for rescanning of ballots with no risk of duplicative counting,
- Maintain ballot secrecy,
- Secure the chain of custody from the local precinct to the state,
- Reduce risk of insider attacks, and
- Prevent hackers from changing results.

We implemented the POC on Gravic-owned scanning equipment, with our back-end processing on servers located at the Hewlett Packard Enterprise Customer Experience Center in Alpharetta, GA. We have processed over 22 million test ballots through the POC so far. The POC will be available for demonstration for a limited period of time upon request.

About Remark Innovations

Remark Innovations Inc. is a wholly owned subsidiary of Gravic, Inc., based in Malvern, PA USA. This company is dedicated to *Innovative Solutions for Society*.

Gravic, Inc. has been a world leader in providing innovative data collection, transformation, and distribution solutions for over 40 years. Our software product groups have produced technologically advanced solutions that improve the businesses and personal lives of our over 100,000 customers and tens of thousands of OEM end-user licensees. The company is 100% owned by U.S. citizens.

Contact the authors at information@remarkinnovations.com.