Election Security in Allegheny County and the Commonwealth of Pennsylvania

William R. Cunha, Mary J. Emanuel, Koji Ina, and Salvador A. Velazquez
Heinz College of Information Systems and Public Policy
Carnegie Mellon University
Pittsburgh, PA

May 10, 2018

Abstract Following the revelations of the 2016 Presidential election, election security has been a hot-button issue throughout the country and at all levels of government. This paper aims to contribute towards these efforts by focusing on Allegheny County and the Commonwealth of Pennsylvania. Through targeted research and analysis, we have not only been able to discover important data points that can help decision makers focus their efforts and budgets on the security of election systems and processes, but also provide overall risk analysis for their entire election ecosystem. By focusing on the swing state of Pennsylvania and, in more detail, Allegheny County, our methodology can be extrapolated to a higher level for a more wide-reaching impact. With an initial objective of determining the vulnerabilities, attack paths, and associated risks within the election system of Allegheny County, our results provide an overall risk analysis for leaders of Allegheny County and the Commonwealth of Pennsylvania. This paper encapsulates the entirety of our efforts, including introduction, methodology, results, and recommendations/further areas of study. Our hope is that by helping to move the needle in favor of the security of US elections, the sanctity of the right for all citizens to vote will be fostered and further defended. It will be a long and arduous battle, but one which must be won.

***Index Terms* — Cyberspace, US elections, election security, election risk assessment, government, risk analysis**

I. Introduction

EVERY election throughout the United States, whether it is at the local, state, or national level, embodies one of the most fundamental rights being a citizen of this country holds - the right to vote. With the ubiquitous rise of technology and connected systems, the nation's election infrastructure, recently deemed Critical Infrastructure by Department of Homeland Security Secretary Jeh Johnson, has been the target of attacks by actors of all types, from the nation state and advanced persistent threat (APT), down to the most amateur computer user [1]. This threat is not going away and, as such, must be dealt with swiftly and thoroughly in order to preserve the sanctity of the election process and the nation's voters.

The Presidential election of 2016, although unique for many reasons, offers important similarities that we utilized during our analysis. For example, like Presidential elections in the past, there were key "swing states" which helped shift the pendulum in favor of the victorious candidate. In 2016, these states were the Commonwealth of Pennsylvania, Michigan, and Wisconsin [2]. With the winning candidate winning each state by 0.7% (44,292), 0.2% (10,702), and 0.7% (22,748) respectively, a slight shift of these numbers would have shifted the balance of victory in favor of the losing candidate [3] [4]. Although looking at the entire election, and election process, throughout the country is an important national security mission, the scope and breadth of that endeavor was larger than the time and resources our team had available. As we looked to analyze the security of the election ecosystem, we decided to scope down our analysis to the Commonwealth of Pennsylvania and, more specifically, Allegheny County. The Commonwealth of Pennsylvania was chosen for a few reasons, but primarily because: 1) Pennsylvania was deemed a key swing state in the Presidential election of 2016; 2) Pennsylvania and 48 other states, including the District of Columbia, is a "winner-takes-all" state, with the winner of "the plurality of the popular vote" earning all of the Electoral votes for that state [5]; and 3) Pennsylvania was one of the 21 states notified by the Department of Homeland Security (DHS) as a state specifically targeted by nefarious actors during the 2016 Presidential election [6]. Allegheny County was chosen for a multitude of reasons, but primarily because: 1) Allegheny County is the 2nd most populated county in Pennsylvania [7]; 2) there is a great deal of information and data available about election results in the County (and it is easily accessible); and 3) the location of Carnegie Mellon University within the County allowed us access to influential people with a wealth of knowledge about the process, such as Mr. Ron Bandes of VoteAllegheny.org.

Allegheny County, in addition to being the second largest county in Pennsylvania, is also a county which uses election systems called Direct-Recording Electronics (DREs). Although there are an array of models and types of these systems, the ones within Allegheny County are ES&S iVotronic DREs, which lack a paper-auditable trail of the ballots that have been cast. This feature of DREs is

sometimes referred to as a voter-verified paper audit trail, or VVPAT. Systems with VVPATs provide election officials the ability to audit ballots that are cast in order to ensure the ballot counts of individual machines match the electronic results produced by the machines. Although there remain security concerns with the VVPAT option, the idea of providing a paper-trail for votes that were cast is a notion that is widely supported at all levels of leadership familiar with the election process [8] [9]. With Governor Tom Wolf's recent announcement that Pennsylvania will make a concerted effort to "replace...electronic voting systems with machines that leave a verifiable paper trail by the end of 2019," this is a step in the right direction [10]. However, as it stands, Allegheny County remains at the mercy of DREs being audited manually, with the integrity of the entire process, and every ballot cast, being preserved by the electronic systems being used and the election officials at the state, county and precinct level. Although at times this might help provide some resiliency within the process (e.g. the absentee ballot process), the current way of conducting elections, and the systems being used in Allegheny County specifically, introduces a number of potential attack vectors for willing and able actors.

II. METHODOLOGY

When looking at an ecosystem as expansive as the election process in the United States, there are certain aspects which must be understood before a deeper dive of the shortfalls and potential recommendations for improvement can be offered. As we began this study, our first step was to look at the entire election ecosystem in order to determine which parts of the process we specifically wanted to focus. In the wake of the 2016 Presidential elections, there has been a lot of discussion surrounding the security of the electronic voting systems used throughout the country, but we thought focusing on just the election systems would lead to a myopic view of the larger question of election security. Any election is comprised of multiple aspects, involving everything from the registration of a potential voter to the recruitment of election officials, publicity of verified candidates, the casting of ballots, and the aggregation, counting, and auditing of ballots. Instead of focusing solely on the electronic voting systems, we endeavored on the task of understanding the entire process, as it applied to Allegheny County, and then worked towards analyzing different vulnerabilities inherent within the process.

In addition, we identified associated attack scenarios attached to each part of the process and attempted to expound upon what those attacks could potentially yield. With every locality throughout the United States determining its own election processes and procedures, focusing on Allegheny County required us to understand exactly how the process worked locally. After the process was understood, we could then work towards identifying potential vulnerabilities and security gaps, followed by recommendations for improvement and further study.

The question of risk analysis is something that is widely discussed throughout the information security domain, with a vast array of adopted models, frameworks, and associated de facto standards. As we looked to determine how we could specifically quantify the risks posed by certain vulnerabilities within the Allegheny County election processes, we reviewed the currently published literature on widely-accepted risk assessments and methodologies. Reviewing everything from the International Organization for Standardization (ISO) 27001 information security management system to the Factor Analysis of Information Risk (FAIR) and the National Institute of Standards and Technology's (NIST) Risk Management Framework (RMF), we found it difficult to single handedly apply any of these standards to our risk analysis of the election processes within Allegheny County [11] [12] [13].

When looking at these frameworks, a common shortfall we encountered was that risk was generally associated with an equitable monetary value attached to the potential loss. Although this might work for corporations, governments, or other organizations attempting to quantify and prioritize risk calculations as associated with asset loss and the bottom dollar, analyzing risks inherent within the election processes results in a different value attached to the specific risk. Risk in the sense of elections is tied to an unquantifiable result of a ballot being lost, altered, or affected in some other manner. As such, an argument for the association of risk with voter confidence naturally ensues. Does voter confidence increase voter turnout? These, and other similar questions, remain up for debate depending on the context of the situation, but it is clear that election security risk does not equate to a definitive dollar estimate [14].

We utilized the ISO 27001 framework as a starting point for our risk assessment, but modified it in a way that better captured the risk associated with election security. Building on the traditional notion of $Risk = Likelihood * Consequence$, we then expanded the variable of "likelihood" to be determined by the following: 1) privileges needed (type of actors); and 2) the number of actors needed. Privileges needed involves the level of access required in order to affect a certain outcome to a given attack scenario. For example, a voter has a different level of access to the infrastructure and material of a given election than a vetted election official might have. The number of actors is associated with "how many people are required to have this specific effect pertaining to this specific scenario." If a

certain scenario (e.g. modifying an election device) only requires one person, that associated risk is different from a scenario requiring the coordinated efforts of five or more nefarious actors. Furthermore, the "consequence" variable was normalized to represent a value we determined to be consequential and meaningful for attributing a notable sense of risk within a given attack scenario, and thus an election. Further details of our risk assessment methodology and calculations follow within the "Data Analysis" and "Risk Assessment" sections of this paper.

III. DATA ANALYSIS

The objective of our data analysis was to determine a percentage threshold representing the number of votes required to be compromised in order to flip a given Pennsylvania election in favor of the second place. The analysis utilized the election turnouts of all 56 elected offices in Pennsylvania in general elections held between 2000 and 2016. The analysis included the following State and Federal offices: Attorney General, Auditor General, Governor, President, State Representative, State Senator, State Treasurer, U.S. Representative, and U.S. Senator. Reference [15] contains the datasets for all 56 elections we utilized for our data analysis.

The analysis encompassed data collection, cleaning, aggregation, and visualization. The collection of data consisted of processing the data available in the repository. The cleaning process consisted of eliminating observations where the z-value for the difference between the first and second place candidate was over |3| standard deviations from the mean. The aggregation of the data consisted of grouping election results by the range of percentage thresholds identified in the analysis. Finally, the visualization consisted of visually representing the results of the data analysis, primarily via a heat map displaying results of the aggregation.

The scenario that we utilized throughout the data analysis portion was the specific case of when an attack compromises votes to favor the second place. Therefore, each of the 56 elections has a critical value representing the number of votes needed to be changed in favor of the second place in order to make the second place candidate win the election. Although we began our data analysis in search of a specific critical value that would represent the tipping point of any given election, we realized that a percentage critical value would better allow us to compare proportions of total votes across all 56 elections and better measure potential impact within each respective election. With this framework in place, we found that given less than 0.6% of the total votes being compromised, no attack would have changed the results of any election between 2000 and 2016. By contrast, if more than 22% of the total votes cast would have been compromised, all 56 elections would have been flipped in favor of the second place candidate. Therefore, our range for critical percentage thresholds is from 0.6% to 22%, as displayed in Figure 1 of Appendix I.

With this information, we chose 10% as the percentage threshold to be used for the individual attack scenario's risk analysis. With 10% of the total votes being flipped in favor of the second place candidate for all 56 analyzed elections, 38 out of the 56 (68%) elections from 2000 to 2016 would have been flipped in favor of the second place candidate. Furthermore, according to our model, if the total turnout in the 2020 Presidential election in Pennsylvania were 5 million votes, an attack favoring the second place candidate would need to change 500,000 votes in order to have a 68% chance of making the second place candidate win the election. It is important to note that the 10% threshold of compromised votes can be originated from an attack within any Pennsylvania county and does not have to reside solely within Allegheny County. Consequently, using the same example and methodology, the alleged 500,000 compromised votes could be accomplished via an attack changing 80% of the total turnout in Allegheny County (about 600,000 votes) alone, or by changing 10% of the total turnout throughout all 67 counties in Pennsylvania.

It is important to note that our model is assuming that all 56 elected offices from 2000 to 2016 are comparable and weigh the same. This is an important supposition because, for the sake of the model, we assume State and Federal elections are comparable and that the consequences of changing the result of a Presidential election are the same as the consequences of changing, for example, the result of an Auditor General election.

IV. ELECTION PROCESSES

Part of the difficulty of analyzing the election process is that the election processes throughout the United States are quite different from one another. Each state regulates itself, with not only states varying in how elections are conducted, but counties and precincts varying in processes and standards as well. Voting on different sides of the same state could result in vastly different experiences. This variety is partially what makes the discussion surrounding the improvement of the security of elections so difficult. The National Institute of Standards and Technology (NIST) and the Election Assistance Commission (EAC) are currently in the process of updating the Voluntary Voting System Guidelines (VVSG), with a large part of their work describing the election processes in broad enough terms so that they can be applied to all stakeholders [16]. Similarly, one of the first tasks we undertook was using the VVSG working group models to map Allegheny County's processes specifically. Figure 2 of Appendix I shows the seven uses cases that are

explained below: Voter Registration, Electronic Poll Book, Ballot Delivery, Ballot on Demand, Ballot Marking, Election Night Reporting, and Auditing.

A. Voter Registration

Voter registration is the process eligible voters must undergo before they cast their ballot in an upcoming election. This can be done via mail, in person at participating government offices (like Pennsylvania's Department of Transportation or the County Assistance Office), or online [17]. There were 8,646,238 registered voters in Pennsylvania in 2016 [17]. In 2016 alone, 915,691 people registered to vote online and 456,820 registered via paper [17]. If submitted via paper, the forms are sent directly to the Allegheny Elections Division to be processed. However, online registrants are first processed by the Pennsylvania Department of State and then sent to the County's election division office. The Allegheny Elections Division Office validates registration applications and enters the approved registrants into the Statewide Uniform Registry of Electors (SURE) database. The SURE database is used by all counties in Pennsylvania to keep a centralized system of voter registration, maintain voter history, and aid in generating absentee ballots, poll books, and election reports [17]. The SURE database is also used by voters to check registration status and locate their polling place [17].

B. E-Poll Books

A poll book is the list of registered voters that is used on Election Day during check-in. While some counties have transitioned to using electronic poll book software that integrates with other election management software, Allegheny County still uses paper poll books that are printed from the SURE database. The poll book consists of two parts: (1) the voter certificates, which are individual pieces of paper printed for and signed by each voter during check-in; and (2) the district register, which is a binder of each registrant, their information, and signature, which is used for comparison to the signature on the voter certificate to validate the voter's identity [18].

C. Ballot Delivery

Ballot delivery refers to both the process of transporting a blank ballot to the registered voter and transporting a marked ballot to a location for processing. The majority of voters do not come into contact with this use case because their ballot is cast at the polling station, in person. In Allegheny County, the ballot delivery process is rather straightforward. Primarily, there are two types of ballots which can be applied for prior to the election, absentee ballots or alternative ballots. Absentee ballots require the voter to establish that they meet a number of requirements, but primarily they must be a registered voter and be:

- Unable to attend polling place due to illness or physical disability; or
- Absent from municipality of residence on the day of the election during the time the polls are open due to duties, occupation, or business; or
- Not attending polling place on the day of the election during the time the polls are open due to observance of a religious holiday [19].

Registered voters must apply for a blank absentee ballot by "5:00 PM on the Tuesday before the Primary or General Election" and can do so via an online system, in person, or in writing" [19]. Once the application is received, a blank ballot will then be sent for the applicable elections being requested or upcoming (depending on if it is a primary or general election). The 2016 general election had 37,050 absentee ballots cast in Pennsylvania [20]. Alternative ballots operate in relatively the same fashion, but the requirements are slightly different. In order to be eligible for an alternative ballot, the person must be a registered elderly or disabled voter and assigned to an inaccessible voting location. Furthermore, the voter must be "65 years of age or older, or have a temporary or permanent physical disability which prohibits you from entering a polling place with architectural barriers; and the polling place has been designated as 'inaccessible'" [21]. Although a current number was unavailable at the time this report was written, Mr. Ron Bandes estimates the number of "inaccessible" polling places in Allegheny County to be around 2-3 out of 1,322 as of 2016. After they are filled out, both absentee and alternative ballots are returned to the Division of Elections primarily via the United States Postal Service (USPS). These ballots can also either be brought to the Division of Elections or the voters' applicable precinct, in person. Once received by the Division of Elections, the ballots are sent to the seven different regional centers in order to be provided to the Judge of Elections for each precinct, where they are then taken to the precinct for election night. According to Mr. Ron Bandes, a Judge of Election in Pennsylvania, both absentee ballots and alternative ballots are kept on-site during the election for three primary reasons: (1) in case the voter shows up in person and the absentee ballot needs to be voided; (2) to be able to set the ballot aside if it is challenged by a poll worker or poll watcher; and (3) to produce an unofficial tally of absentee ballots at the closing of the precinct. Once the precinct is closed, the ballots are transferred to the Division of Elections where they are counted by the ES&S Model 650 high-speed optical scanners [8], after which point the results are transferred to the Unity system for official tabulations.

D. Ballot on Demand

Ballot on demand refers to the type of ballot generation used by states that have alternative voting processes such as

creating blank ballots at the polling location or at home. An example would be a state that allows voters to vote at any polling location, regardless of assignment. This process facilitates the generation of ballots wherever the voter chooses to vote. There are no instances of ballot on demand in Allegheny County.

E. Ballot Marking

Ballot marking is the actual process of the voter casting their vote – either by checking boxes on a paper ballot or by using a voting machine to select their candidates. Since we have already reviewed the paper process in Allegheny County via the ballot delivery use case, this will focus primarily on the electronic voting machines. The DRE system that Allegheny County has certified for use is the ES&S iVotronic 9.1.4.1, with 4,508 in use throughout the county [8].

Once the voter has been verified as registered, the poll worker will use a Personal Electronic Ballot (PEB) to prepare the DRE for that specified individual's vote. The PEB communicates with the DRE via infrared communications, which is a technology that has been around for quite some time. Once the voter has cast their vote using the touchscreen, the vote is stored both in the internal flash memory of the DRE, a compact flash (CF) card, and the PEB itself [8]. The summary results from each PEB at the polling location are printed and used for the counting and auditing process later on [18].

F. Election-night Reporting

Election-night reporting is how votes are reported to the media and the public. The immediate results printed from the PEBs of each voting location are posted on the door of the polling location after voting has closed. Poll workers collect the PEBs, CF cards, printed results, and paper ballots from each precinct into a single packet, which are then transported by the Judge of Elections to one of the seven regional centers for the county [18]. Votes are tabulated by the Unity software at the regional center and transferred via modem landline to the Election Division's warehouse. The master PEBs are read by the Unity system in order to get the official tabulation of votes, which may be supported by reading the CF cards, if necessary [22].

After the voting materials are transferred and counted at the Election Division warehouse, the unofficial results are uploaded onto the publically-facing web portal. For the 2016 general election, unofficial results were finalized by November 10, 2016 and certified results were posted on December 12, 2016 [23].

Increasingly, media companies will use third-party data collected from front-line "stringers" and exit polls to model election results in real time. The raw information is vetted through "sanity-checks" to ensure models are staying within pre-determined and calculated expectations (usually determined by proprietary algorithms and experienced election analysts). These mechanisms allow for media outlets to provide faster predictions and help inform the public about the election results before the unofficial counts are released.

G. Auditing

Auditing is the post-election process of verifying the integrity of the election results. Beforehand, however, there are a number of tests performed to ensure the equipment is functioning properly. Two months before the election, a Logic and Accuracy (L&A) test is performed by a third-party on 20 randomly-selected DREs to ensure the firmware has not been altered [22]. On the day of the election, two DREs and 1 PEB are picked to undergo parallel testing to simulate election conditions in order to see if the output matches the input [20]. The flash cards from the DREs are used to validate the numbers from the master PEBs in Allegheny County [22]. The election tabulation network is also reviewed by third-party consultants to ensure the network is separated from outside devices [20] and safely able to tabulate and store election results securely.

V. Allegheny County Attack Scenarios

After understanding more about the election processes for Allegheny County, the next step for us was to determine vulnerabilities that could allow malicious actors to influence the election in some way. In order to stay in line with the assumptions of our methodology - that a compromised vote would go from the first place candidate to the second place candidate - we only focused on scenarios that could have a measurable effect in this manner. We identified six scenarios that fit our criteria: (1) an attack on the online voter registration system; (2) an attack on marked absentee ballots; (3) an attack on the DREs from the Pennsylvania Election Division; (4) an attack on PEBs in the Pennsylvania Election Division; (5) an attack on DREs using the PEB in the precincts; and (6) an attack on the Unity system using a malicious PEB. An overview of the election processes coupled with the associated attack scenarios is outlined in Figure 3 of Appendix I.

A. Attack Online Voter Registration Form

Online voter registration became available to Pennsylvania residents on August 25, 2015 [17]. Since its introduction, the online voter registration portal has become an increasingly popular way for applicants to send in their voter registration. However, the weak authentication required of the applicants sending in registration forms is a major vulnerability to this process. In order to submit a form on the online voter registration portal, an actor only needs to submit the name, current address, and proof of identify. This proof can be

either a Pennsylvania driver's license or the last four digits of their social security number – both of which are available from data breaches on sites like Pastebin or for purchase on the dark web.

While numerous fake registrants may be detected in the database, attackers can also modify information of existing registrations with the option to change someone's name, address, or party. A malicious actor could abuse this ability in order to move voters to different polling locations that already lean in the direction of their party or to locations the voter would be unable to reach on short notice. These alterations would deem a valid voter ineligible to vote in an assigned polling location. Election officials may be able to contact the Elections Division Office on Election Day to determine a voter's correct location, but there could be issues with this information being determined in time or with the voter traveling to the alternative location [18]. The voter would be allowed to cast a provisional ballot if no information was verified, but these votes are not counted until the Friday after the election and face scrutiny by a County review board [18]. An educated attacker would be able to pinpoint which voters to target in order to influence the outcome of the election in the county.

With the rise of personal information databases for sale on the dark web, it has been shown to be relatively easy and inexpensive to obtain personally-identifiable information. A recent Harvard study found that it would cost an attacker, on average, $31 to obtain the necessary information about a given voter and automate the attack on the voter database to change 1% of the total votes (87,223) and $315 to change 10% (872,228) in Pennsylvania [24]. Although this study was focusing on attacks at the state level, it could be possible to perform the same attack for the same cost only on Allegheny County. Because the state database with a list of voter names, addresses, precinct, political party, and voter history can be purchased legally from the Pennsylvania SURE web portal for $20 for the entire state, the only modification to narrow the scope of the attack at the county level would be to focus which records of licenses and SSNs are found or purchased. The polling location tool also made available by the state of Pennsylvania would further aid malicious actors by determining the polling location of each voter from their address in the database [25].

Although it is possible to target users and change their registration information, scaling up an attack to have a measurable impact could prove to be a challenge, depending on how much resources an attacker or hacker group has at their disposal. Automation, as explained, could make this process much easier. Targeting users could be done by sorting by their polling location using the tool provided by Pennsylvania or by obtaining the latitude and longitude of each address and determining polling locations from a separate dataset. One journalist already found evidence that others may be working on similar efforts. Jonathan Albright, the research director of Columbia University's Tow Center for Digital Journalism, found a project on GitHub that references voter ID and Congressional District in a script that is supposed to find geographical coordinates based on address, which could also be altered to find new addresses to assign targeted users outside of their polling location [26]. What is more concerning is that the GitHub account hosting this information belonged to a data science intern from Cambridge Analytica named Michael Phillips [26]. The inclusion of voter IDs and assigned congressional districts suggests that the code was at least a proof of concept for manipulating voter identity information in some way.

Even more troubling is that another script on Mr. Albright's GitHub account, titled "Twitteranalysis.py," essentially used sentiment analysis to determine how users felt about certain political issues and candidates [26]. Although there is no proof that the two scripts were used together, if they were combined with voter database information, they could be used to identify real-time sentiments of actual voters, determine which individuals to target based on their political beliefs, identify their address, and reassign them to a different location [26]. Depending on whether the Pennsylvania Elections Office logs the IP address of changes to the voter registration database, it would be very hard to track down which changes were legitimate and which were from nefarious attackers.

B. Attack Marked Absentee Ballots

For both the absentee and alternative ballots, once a voter has requested the blank ballots, they are then sent to the voter via the United States Postal Service, without any type of extra protection and security provided (i.e. no Certified or Registered Mail services). Once the voter has received the ballot, he or she will mark the ballot as they desire and place the ballot in an unmarked envelope, which is then enclosed inside a pre-paid mailing envelope and placed in the mail for delivery to the Division of Elections in Allegheny County. This mail is transferred in the same manner as the blank ballot, without any additional proof or guarantee of delivery. This is the primary attack method outlined in this section. Should the handler of the mail have nefarious intentions or just perform their duties in a less than satisfactory manner, the sanctity of the voter's ballot could be either altered, compromised, or lost. Without any type of control placed on the ballot delivery process, from the voter to the Division of Elections, no auditing or security is offered to the voter to ensure their ballot was cast as originally intended, or received to begin with. As such, this attack vector, although low in potential reach (only 6% of ballots throughout Pennsylvania were cast via absentee or

alternative in 2016), remains a viable attack scenario.

C. Attack DREs in PA Election Division

The third attack we identified is a modification of the firmware of the DREs. Although the firmware of these machines are examined before the election, only twenty machines are examined two months before the election [20]. Additionally, a logical and accuracy test is performed on all DREs a month before the election [27]. However, an attack could still be carried out between the final test and the day of election. Research from the EVERST project found that the firmware of the iVotronic DRE could be altered using a screwdriver and a standard EEPROM programmer to change votes cast for one candidate to be marked and counted for another candidate [28]. In 2016 there were 620,552 votes cast on 4,508 DREs, which means each machine averages about 136 votes during the election [23]. In order to change a 10% gap between one candidate and the other, at least 500 DREs would need to be compromised in this manner.

Although the County does run a parallel test on Election Day by simulating voting inputs to ensure the votes are counted in the way they are cast, this integrity test only applies to two randomly selected machines at the same time the election is going on [20]. If 500 DREs were successfully compromised to carry out this attack, there is only a 20% chance that one of those machines would be selected for the integrity test. Finally, the actor could revert the firmware back to its original state before the next tests are done to the DREs. Because of the necessity to have access to the DREs before and after the election, it is most likely that the actor would need to have privileges to the Elections Division Warehouse, either as an insider threat or with the help of an employee.

D. Attack PEBs in Election Division

The fourth scenario we envisioned is to attack the PEBs in the warehouse to compromise the DREs when they are used by the poll workers. On Election Day, the Election Division staffers load the ballots for each precinct from the Unity software onto the PEBs to be used at the polling location. Election staffers copy ballots from the PEB to the DRE in order for each voter to cast their ballot. By compromising PEBs ahead of time, an attacker could modify the DREs used during the election process. An attacker would need to compromise 32 DREs in order to close the 10% gap between the first and second place candidate.

The most important part of this attack scenario is the existence of a buffer overflow vulnerability in the DRE model used by Allegheny County. PEBs can exploit this weakness by inserting specially crafted large strings in the ballot definition [28]. Once the buffer overflow is exploited, the DRE firmware could be modified in the same way as in the previous attack, resulting in votes being cast for one candidate instead of the intended candidate.

Although network vulnerability tests have shown that the Allegheny County's Election Division Tabulation Network does not connect to the Internet other than through the dial-up modem to regional centers via serial port, there are many vulnerabilities within the network itself [20]. A malicious staffer could load the PEBs into the PEB writer on the network and modify them in the warehouse. If they did not have credentials to the Unity system, the software's authentication process could be bypassed with a SQL attack [28]. It would technically be possible for outsiders to carry out this attack as well – multiple reports have shown the physical doors to the warehouse being left open and the security footage is only retained for a single week [29] [30] [31].

E. Attack DREs by PEB in Precincts

The fifth attack identified is to modify the DREs by using a PEB at the polling location. Because the PEBs and DREs communicate via infrared, important information can be intercepted and utilized to exploit the same buffer overflow outlined in the previous attacks. The supervisor PEB uses a qualifications code and key to authenticate itself to the DRE [28]. This is transmitted in plaintext and could be intercepted, copied, and used by an attacker to authenticate their PEB emulator with any infrared-capable device [28]. From there, the PEB emulator could exploit the DRE's buffer overflow vulnerability and modify the election results. While it would require more effort to get the authentication information for each supervisor PEB on Election Day, an insider in the Election Division could also obtain this information ahead of time. An insider could also use a compromised PEB or PEB emulator on the day of the election to carry out this attack more easily than a third-party.

This attack would need to compromise more PEBs than attack scenario 4 because there is a difference between modifying intercepted votes and loading marked ballots in the DREs. Attack scenario 4 would load marked ballots in the DREs, which means that it would be able to change any votes cast (for 1st, 3rd, etc. candidates) to the second candidate. It would also be able to mark ballots that are not actually cast if a registered voter does not show up. Attack scenario 5 only intercepts votes in the DRE – this means that it can only change actually marked ballots, not ballots that go unused. Because of this difference, the number of DREs needed to bridge the 10% gap for attack scenario 4 is 32, while the number of DREs needed for attack scenario 5 is 50.

F. Attack Unity System via PEB Emulator

The sixth attack scenario would be to compromise the Unity software used to manage the election via a malicious PEB. After the polls are closed, the tabulation of the ballots

is performed by the Unity system in a central location by reading the PEBs and flash memory from the DREs [18]. A specifically crafted PEB could exploit a stack-based overflow in the Elections Reporting Manager to gain control of the entire Unity system, which could allow an attacker to modify the tabulation and auditing results [28]. This means that only one malicious PEB would be needed to compromise 10% of the votes. However, if any of the paper results printed at the polling location were used to audit the election, the attack would easily be detected.

VI. Risk Assessment

Determining a proper calculation for risk in these scenarios was difficult because many of the risk management frameworks in use today focus on the quantitative amount of money lost as the measurement of risk. However, the dollar amount loss associated with each attack path is less important than upholding the integrity of the election and individual ballots. For this reason, we created our own risk table based off of the general understanding of risk and the principles from ISO 27001. The calculation we used for determining risk was the following:

$$Risk = Normalized\ Consequence * [Privileges\ Needed * Number\ of\ Actors\ Needed]$$

Because we were interested in attacks that could close the 10% gap between the first two candidates in an election, we normalized the consequences for all attack scenarios to a 10% impact. We defined likelihood as being the number of actors needed and the privileges of each actor needed to carry out the attack. The results of our risk assessment are contained in Figure 4 of Appendix I.

A. Attack Online Voter Registration Form

The voter registration form attack requires no special privileges or access. It could be carried out over a long period of time or utilize automation, so it only needs one actor to be successful. Because it requires no special access and a low number of actors, it has a high risk score of 9.

$$Risk = (1) * [3 * 3] = 9\ (High)$$

B. Attack Marked Absentee Ballots

Attacking marked absentee ballots would have to be done by a postal worker in transit or an Elections Division worker at the point of delivery. However, as only 6% of votes in Pennsylvania are voted absentee, it would not be possible to reach the 10% gap of votes needed for the expected consequence. Additionally, the challenges of intercepting such a large number of votes would make it quite difficult to carry out this attack scenario.

C. Attack DREs in PA Election Division

In order to carry out an attack on the DREs in the Election Division warehouse, the actor would need a high level of privileges, such as from an election official, and around 5-10 actors would be needed to compromise 500 DREs in the short time frame of one week. For these reasons, this attack scenario has a low risk score of 2.

$$Risk = (1) * [2 * 1] = 2\ (Low)$$

D. Attack PEBs in PA Election Division

Attacking the 32 PEBs in the Election Division warehouse to later compromise the DREs would require a high level of privileges, such as from an election official, but would only require 1-4 actors to be successful. These factors give this scenario a medium score of 3.

$$Risk = (1) * [1 * 3] = 3\ (Medium)$$

E. Attack DREs by PEB in Precincts

To attack the DREs through the PEBs at the polling locations, at least 4 supervisor PEBs would need to be compromised. This attack would be most successful if the qualification codes and keys were obtained beforehand, or if the poll workers used the compromised PEBs or PEB emulators to carry out the attack. There would need to be over 10 actors to be successful on Election Day, which puts this scenario at a low risk score of 2.

$$Risk = (1) * [2 * 1] = 2\ (Low)$$

F. Attack Unity System via PEB

Compromising the Unity software via a PEB would require the privileges of a poll worker to craft the malicious PEB before the tabulation process. However, it would only require impacting one PEB, giving it the high risk score of 6.

$$Risk = (1) * [2 * 3] = 6\ (High)$$

Determining the amount of privileges needed and number of actors needed to carry out each attack allowed us to determine which attack scenarios were more likely to be utilized by malicious actors on Allegheny County. Knowing the likelihood of each attack also facilitates the prioritization of which vulnerabilities should be addressed first in order to protect the integrity of the election process for the county.

VII. Conclusions

A. Remediation for Attack Scenarios

This section outlines our recommended remediation for the previously described attack scenarios.

1) Remediation for Attack on Online Registration Form

The main vulnerability for the online voter registration attack is the weak authentication required to

make changes to the voter database. Even requiring a link to be clicked from an email associated with each account (i.e. two-factor authentication) would improve security and hinder voter records from being changed on a mass scale. Additionally, we did not find any information about logging practices for the portal. If this is not already done, collecting information about time stamps and IPs could provide some investigatory evidence about if the registration form is being abused.

2) Remediation for Attack on Marked Absentee Ballots

Although attacking marked absentee ballots was not significant according to our threshold, it could still be exploited with other attacks. We recommend investigating methods to provide extra security for the ballots in transit, such as USPS Certified Mail, to ensure a safe chain of custody of mail-in ballots.

3) Remediation for Attack on DREs in PA Election Division

To prevent DREs from being modified in the PA Election Division Warehouse, there should be more surveillance cameras installed and the recording cameras should store a longer period of activity. The physical security of the warehouse should also be improved. We recommend keeping the parallel tests on the day of the election, but think, there should be more DREs in the test group to have a higher rate of detection. For example, having six randomly selected DREs instead of two would increase the rate of detection from 20% to 50%.

4) Remediation for Attack on PEBs in PA Election Division

Implementing random checks of the PEB's firmware immediately following the election would detect maliciously crafted PEBs. The publicly known vulnerabilities should be remediated, either by installing certified patches from the vendor or switching equipment all together.

5) Remediation for Attack on DREs by PEB in Precincts

Because qualification codes and encryption keys can be extracted via infrared communications at the polling location, there should be a higher emphasis placed on the importance of securing the PEBs during poll worker training to keep them from being accessed by voters.

6) Remediation for Attack on Unity System via PEB

As one malicious PEB could compromise the entire ballot tabulation component of the Unity system, paper results should be used in the auditing process. In addition, the buffer overflow vulnerability should be remediated with patches or replacement.

B. Recommendations to Allegheny County Elections

Between now and the next election, we have five suggestions to immediately improve the overall security of the election process in Allegheny County: (1) the budget should be allocated in a way that corresponds to the risk of these scenarios, giving more money to the controls that are associated with the high-risk attack paths; (2) the weak authentication to make changes to the online voter registration database should be remediated immediately; (3) using the paper print outs from the PEBs during the auditing process would be useful until new voting systems that include paper records are incorporated into the election process; (4) all poll workers should be fully informed about the vulnerability surrounding the PEBs in their precincts and the care that needs to be given to the handling of these sensitive items; and (5) thorough background checks should be done for all individuals who have access to the Division of Elections warehouse and network.

Looking beyond the midterm elections in 2018, there are also long-term changes that need to be made to the process. Incorporating some system of patch management is important for continuing to maintain the security of elections systems. While the State must currently re-certify software before system patches can be applied, changing this constricting legislation to allow thoroughly vetted patches without an additional, bureaucratic certification process would eliminate many technical vulnerabilities more expeditiously. Paper auditing should continue to be made a priority and incorporated in the auditing process. Additionally, adding an emphasis on the principle of least privilege to the policies and procedures to the Elections Division will help alleviate some of the threats from insiders.

C. Lessons Learned

Our team took away four main lessons as a result of our research. The first is that there cannot be too much of an emphasis placed on the importance of evaluating election processes holistically. It does not matter how secure the voting machines are if the voter registration database can be influenced earlier in the process. The most vulnerable parts of the system could be an open door to the warehouse, not a perfectly crafted exploit utilizing a known, or unknown, vulnerability.

This goes hand-in-hand with the fact that there is a large gap between what is possible and what is practical. While there was quite a lot of media hype after the Voting Village at DEFCON compromised all of the voting machines on display, there are very few instances in which an attacker would have complete access to these machines, with all of their tools and the time necessary to conduct these attacks. It is even less likely that an attacker would be able to have complete access to enough machines to have an impact on

the results of the election. For this reason, centralized systems like the SURE database or the Unity system are much more important to prioritize for attention than distributed systems such as the DREs or PEBs that would require compromising much higher numbers of units.

Finally, more work needs to be done at the intersection of vulnerability analysis of election processes and the data analysis of how these attack scenarios can influence the election. We found little research that showed how much impact specific vulnerabilities would have on a given election.

D. Limitations

The limitations of this research can be summed up into three points: (1) lack of data; (2) lack of information; and (3) lack of time. Each state has different formats for their election databases, making it harder to apply our model to all 50 states. The lack of information came from the fact that the entirety of our research came from open-source documents. While we were somewhat successful due to the fact that Allegheny County publishes more documents than most, much of the information could be out of date. Unfortunately, we were unable to obtain enough information about other counties in the time frame allotted to test our model outside of Allegheny County. It took a lot of time to piece together the county-level processes from different sources, which is another reason why our final product focused only on Allegheny County.

E. Suggestions for Future Work

Given our limitations, we have three suggestions for future work. The first is to adapt our methodology into an election security risk framework. The applicability of our methodology needs to be examined by using it in other counties and states. The results of this undertaking could be a communication tool for leadership involved in US elections to discuss election security risk in a uniform way.

The second is that any research in this area should use the Center for Internet Security's Handbook for Elections Infrastructure Security to incorporate the industry's best practices moving forward [32]. Although not completely within the scope of our project due to us evaluating a current system instead of creating a new one, we found the guidance helpful when attempting to understand how to overcome the challenges of election security.

Third, we hope to see a collaboration between the Allegheny County Division of Elections and the Heinz College of Information Systems and Public Policy. Heinz College has an enormous amount of expertise and resources available to it and leveraging those resources, especially including the larger Carnegie Mellon University ecosystem, could bring a great deal of knowledge to leadership to help improve the election processes at the county and state level. A key part of this partnership is to ensure that information is effectively shared across all levels of the organizations in order to ensure the most accurate and up-to-date information is being used for any follow-on projects and work.

F. Closing Thoughts

As our study and this report have demonstrated, the security of our election systems and processes is, and will continue to be, under attack. Although this project was limited to the Commonwealth of Pennsylvania and Allegheny County, this type of study must be repeated across the rest of the country. With Pennsylvania being a demonstrated swing state in previous elections, the attack scenarios, risk assessments, and recommendations must be adopted as quickly and efficiently as possible, taking into account the limited resources that are available to all parties involved.

Recently, the Senate Intelligence Committee released a report highlighting some of the 2016 Election findings, with a wide-array of startling and worrisome comments. As mentioned in the report, "At least 18 states had election systems targeted by Russian-affiliated cyber actors in some fashion [33]." Although we have previously mentioned that the provided attack scenarios could be exploited by a wide-range of actors, these findings indicate a clear intent of highly-skilled and motivated actors (i.e. advanced persistent threats) to meddle in our nation's election process. They go on to state that "in a small number of states, these cyber actors were in a position to, at a minimum, alter or delete voter registration data…" [33]. Furthermore, "Paperless DRE voting machines…are at highest risk for security flaws…" and "potentially vulnerable systems include…systems affiliated with voter registration databases, electronic poll books, vote casting, vote tallying, and unofficial election night reporting…" [33]. These findings further corroborate the attack scenarios we have previously mentioned. Most notably, and as we have described with the case for Allegheny County, the report states that "…the Committee notes that a small number of districts in key states can have a significant impact in a national election" [33].

As for recommendations, the report cites steps such as reinforcing States as the primary lead on conducting elections, improving information sharing across all levels of government, and expediting the security clearance process for state and local officials as important next steps [33]. They also recommend that steps such as "institut[ing] two-factor authentication for state databases…updat[ing] software in voter registration systems…[and] any [voting system] purchased going forward should have a voter-verified paper trail…" as vital steps that should be taken at the local and state level [33].

A holistic review of the election system within the United States, on a state by state level, is long overdue and must be placed at the top of leadership's priority list. Although this report is a good start, it does not go far enough. The Commonwealth of Pennsylvania should utilize our findings to set the standard for the rest of the country on how voting systems and processes should be analyzed, assigned quantifiable risk calculations, and assigned proper controls based on leaderships' decisions. The sanctity of the process, the confidence of the voters, and our national security depend on it.

VIII. Appendix

The following appendix is included for further clarification and reference.

Appendix I - Figures

IX. Acknowledgment

This report represents the culmination of many hours of effort throughout the course of an entire semester of work. The authors of this report and project would be remiss without acknowledging the many individuals who offered their support, expertise, and time to help make this a successful endeavor. We would like to acknowledge and thank the following individuals for making all of this possible: Mr. Brett Tucker and Mr. Dan Kambic with SEI/CERT, our advisor Mrs. Summer Fowler, Mr. Ron Bandes with VoteAllegheny.org, and Mr. Jeremy Bowers with the New York Times. Thank you for all of your help.

References

[1] Department of Homeland Security, "Statement by Secretary Jeh Johnson on the Designation of Election Infrastructure as a Critical Infrastructure Subsector," Department of Homeland Security, 6 January 2017. [Online]. Available: https://www.dhs.gov/news/2017/01/06/statement-secretary-johnson-designation-election-infrastructure-critical. [Accessed 18 April 2018].

[2] T. Meko, D. Lu and L. Gamio, "How Trump won the presidency with razor-thin margins in swing states," 11 November 2016. [Online]. Available: https://www.washingtonpost.com/graphics/politics/2016-election/swing-state-margins/. [Accessed 28 March 2018].

[3] W. R. Mebane, Jr and M. Bernhard, "Voting Technologies, Recount Methods and Votes in Wisconsin and Michiganin 2016," University of Michigan, Ann Arbor, 2016.

[4] PA Department of State, "2016 Presidental Election," PA Department of State, 2016. [Online]. Available: https://www.electionreturns.pa.gov/General/SummaryResults?ElectionID=54&ElectionType=G&IsActive=0. [Accessed 18 April 2018].

[5] National Archives and Records Administration, "Frequently Asked Questions," National Archives and Records Administration, [Online]. Available: https://www.archives.gov/federal-register/electoral-college/faq.html#wtapv. [Accessed 2 May 2018].

[6] S. Horwitz, E. Nakashima and M. Gold, "DHS tells states about Russian hacking during 2016 election," Washington Post, 22 September 2017. [Online]. Available: https://www.washingtonpost.com/world/national-security/dhs-tells-states-about-russian-hacking-during-2016-election/2017/09/22/fd263a2c-9fe2-11e7-8ea1-ed975285475e_story.html?noredirect=on&utm_term=.b13f7e47dcab. [Accessed 18 April 2018].

[7] DataUSA, "Allegheny County," Data USA, [Online]. Available: https://datausa.io/profile/geo/allegheny-county-pa/. [Accessed 2 May 2018].

[8] Verified Voting, "Election Systems and Software (ES&S) iVotronic," 2017. [Online]. Available: https://www.verifiedvoting.org/resources/voting-equipment/ess/ivotronic/.

[9] Z. Shoorbajee, "Spooked by election hacking, states are moving to paper ballots," Cyberscoop, 12 March 2018. [Online]. Available: https://www.cyberscoop.com/paper-ballots-election-security-electronic-voting-machines/. [Accessed 18 April 2018].

[10] Associated Press, "Pa. Governor Wants To Replace Electronic Voting Systems By 2020," KDKA-TV, 12 April 2018. [Online]. Available: http://pittsburgh.cbslocal.com/2018/04/12/pennsylvania-electronic-voting-systems-replaced/. [Accessed 18 April 2018].

[11] ISO, "ISO/IEC 27000 family - Information security management systems," International Organization for Standardization, [Online]. Available: https://www.iso.org/isoiec-27001-information-security.html. [Accessed 18 April 2018].

[12] FAIR Institute, "Fair Institute," Fair Institute, [Online]. Available: https://www.fairinstitute.org/. [Accessed 2 May 2018].

[13] NIST, "Risk Management," NIST Computer Security Resource Center, [Online]. Available: https://csrc.nist.gov/Projects/Risk-Management/Risk-Management-Framework-Quick-Start-Guides. [Accessed 18 April 2018].

[14] MIT Election Data and Science Lab, "Voter Confidence," MIT, [Online]. Available: https://electionlab.mit.edu/research/voter-confidence. [Accessed 18 April 2018].

[15] Open Elections, "Open Elections - PA Repository," Github, 7 January 2018. [Online]. Available: https://github.com/openelections/openelections-data-pa. [Accessed 9 May 2018].

[16] NIST, "Voting Public Working Groups," 23 January 2018. [Online]. Available: https://collaborate.nist.gov/voting/bin/view/Voting/WebHome.

[17] P. A. Cortes, "2016 Report to the General Assembly," The Administration of Voter Registration in Pennsylvania, Harrisburg, 2017.

[18] Allegheny County Elections Division, "Election Officers Reference Manual," Allegheny County Elections Division, Pittsburgh, 2016.

[19] Allegheny County Election Division, "Absentee Ballots," Allegheny County, [Online]. Available: http://www.alleghenycounty.us/elections/absentee-ballots.aspx. [Accessed 18 April 2018].

[20] G. Abramowitz, J. O'Brien, K. Perkoski, L. Szurley, D. Voye and M. Wolosik, "Allegheny County 2016 General Election Experience Report," Allegheny County, Pittsburgh, 2016.

[21] Allegheny County Elections Division, "Alternative Ballots," Allegheny County, [Online]. Available: http://www.alleghenycounty.us/elections/alternative-ballots.aspx. [Accessed 18 April 2018].

[22] C. Wagner, "Performance Audit of the Department of Administrative Services Elections Division for the Period January 1, 2013 through December 31, 2015," Allegheny County Office of the Controller, Pittsburgh, 2016.

[23] Allegheny County Elections Division, "Official Results," 25 March 2018. [Online]. Available: http://results.enr.clarityelections.com/PA/Allegheny/63905/Web02.193333/#/rpt.

[24] L. Sweeney, J. S. Yoo and J. Zang, "Voter Identity Theft: Submitting Changes to Voter Registrations Online to Disrupt Elections," Technology Science, Cambridge, 2017.

[25] Pennsylvania Department of State, "plService," [Online]. Available: www.paportallocator.pa.gov.

[26] J. Albright, "Cambridge Analytica: the Geotargeting and Emotional Data Mining Scripts," 13 October 2017. [Online]. Available:

https://medium.com/tow-center/cambridge-analytica-the-geotargeting-and-emotional-data-mining-scripts-bcc3c428d77f.

[27] M. E. S. Johns, "Vote Allegheny Logic and Accuracy Test Report: Self Test," Vote Allegheny, Pittsburgh, 2016.

[28] P. McDaniel, "EVEREST: Evaluation and Validation of Election-Related Equipment, Standards and Testing," Ohio Secretary of State, Columbus, 2007.

[29] W.-T. Pittsburgh, Director, *Team 4 Investigates Security of Voting Machines.* [Film]. 2010.

[30] W.-T. Pittsburgh, Director, *Team 4: Security Concerns About Voting Machines Remain.* [Film]. 2010.

[31] Solutionary, "Air Gap Assessment Report," Solutionary, Omaha, 2013.

[32] Center for Internet Security, "A Handbook for Elections Infrastructure Security," CIS, New York, 2018.

[33] Senate Intelligence Committee, "Russian Targeting of Election Infrastructure During the 2016 Election: Summary of Initial Findings and Recommendations," 2018.

[34] Allegheny County Elections Division, "Voter Education," [Online]. Available: https://www.alleghenycounty.us/elections/voter-education.aspx.

APPENDIX I – FIGURES

Summation of PA Elections Changed by Percentage of Votes Affected (2000-2016)

Office	0.60%	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%	22%
Attorney General		0	2	2	3	3	4	4	5	5	5	5
Auditor General	0	0	1	2	3	3	3	3	3	4	4	5
Governor	0	0	0	0	0	3	3	3	3	3	3	4
President	0	1	2	4	4	4	5	5	5	5	5	5
State Representative	0	0	3	5	7	8	8	8	9	9	9	9
State Senator	0	0	0	2	4	5	6	8	8	8	9	9
State Treasurer	0	0	1	1	2	3	3	4	4	4	4	4
US Representative	0	3	4	4	4	5	7	8	9	9	9	9
US Senator	0	1	2	2	3	4	5	5	5	6	6	6
Grand Total	0	5	15	22	30	38	44	48	51	53	54	56

Figure 1 - Election Outcomes Heat Map

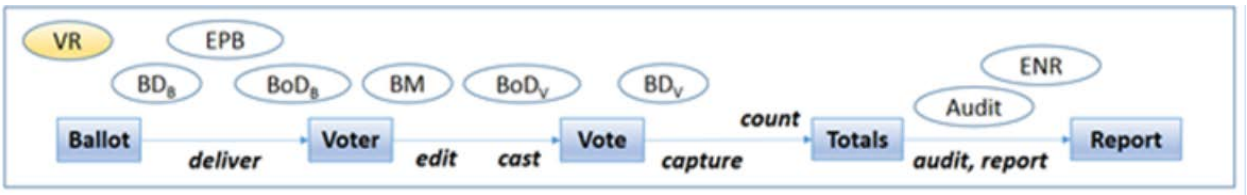

Figure 2 - NIST Election Process Use Cases

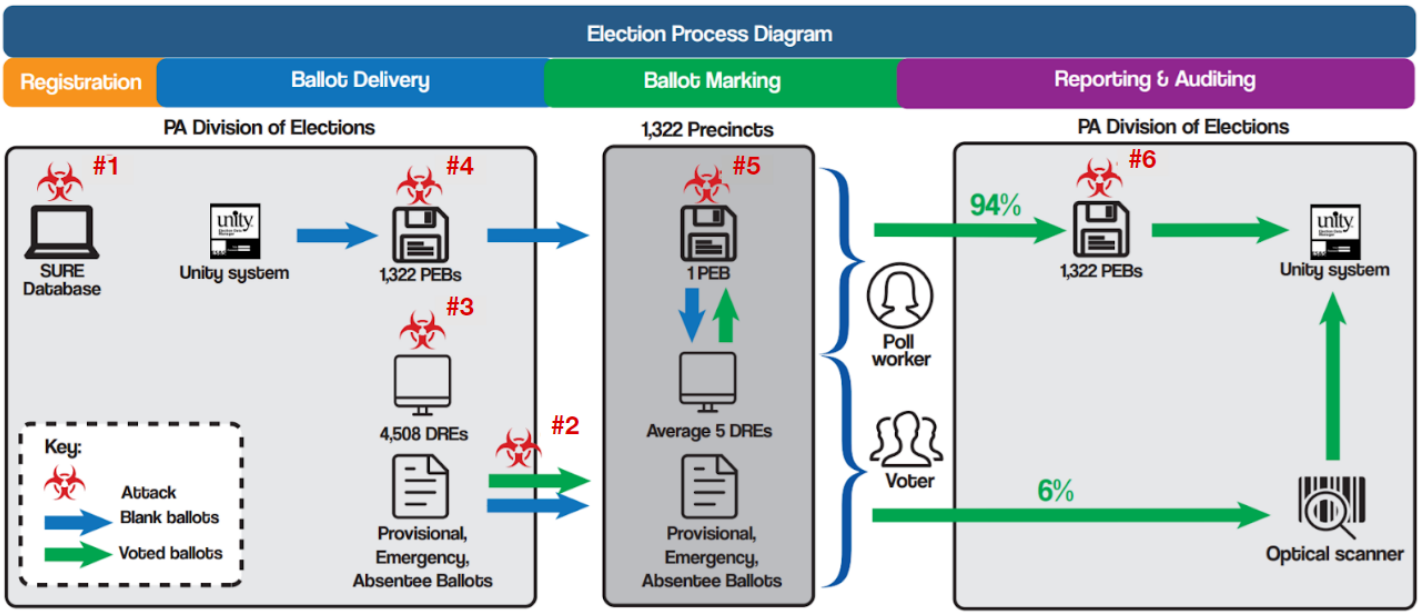

Figure 3 - Election Process Attack Diagram

#	Attack Scenario	Consequence (* normalized)	Likelihood		Risk Score
			Privileges Needed (Type of Actor) High (Election Official) = 1 Medium (Poll Worker) = 2 Low (Voter | Remote) = 3	Number of Actors Needed Unlikely (10 <) = 1 Likely (5 ~ 10) = 2 Very Likely (1 ~ 4) = 3	Low = 1 ~ 2 Medium = 3 ~ 4 High = 5 <
1	Attack online voter registration form	10% < of votes (1 DB)	3	3	9
6	Attack Unity System via PEB	10% < of votes (1 PEB)	2	3	6
4	Attack PEBs in PA Election Division	10% of votes (32 < PEBs)	1	3	3
5	Attack DREs by PEB in precincts	10% of votes (50 < PEBs)	2	1	2
3	Attack DREs in PA Election Division	10% of votes (500 < DREs)	1	2	2
2	Attack Marked Absentee Ballots	5 % of votes	Insignificant consequences to change election result		

Figure 4 – Risk Assessment of Attack Scenarios