# A Mathematical Model of Hacking the 2016 US Presidential Election

**Dennis Nilsson Sjöström**
**Simovits Consulting, Sweden**
**Department of Physics, Umeå University, Sweden**
dennis.nilsson-sjostrom@simovits.com

**Abstract:** After the 2016 US presidential election, allegations were published that the electronic voting machines used throughout the US could have been manipulated. These claims arose due to the reported attacks by Department of Homeland Security toward voter registration databases. The US is more vulnerable against these types of attacks since electronic voting machines are the most prevalent method for voting. In order to reduce election costs, other countries are also considering the replacement of paper ballots with electronic voting machines. This, however, imposes a risk. By attacking the electronic voting machines, an attacker could change the outcome of an election. A well-executed attack would be designed to be highly successful, but at the same time the risk for detection would be low. The question evaluated in this paper is whether such an attack would be possible and if so, how much it would cost to execute. This paper presents a mathematical model of the 2016 US presidential election. The model is based on voting machine equipment data and polling data. The model is used to simulate how rational attackers would maximize their effect on the election and minimize their effort by hacking voting machines. By using polls, it was possible to determine the effort needed to change the outcome of the 2016 US presidential election and thus estimate the costs. This kind of model can be implemented on the elections of other countries that use electronic voting machines. Based on the model, the estimated cost to hack the 2016 US presidential election would amount to at least ten million dollars. The results show that these attacks are possible by attacking only one manufacturer of electronic voting machines. Hence, the use of electronic voting machines poses too much of a risk for democracy, and paper ballots should still be considered for elections.

**Keywords:** critical infrastructure, cyber warfare, cyber weapons, voting security, electronic voting

## 1. Introduction

From the perspective of society, cyber warfare is a rapidly developing new form of warfare. With a larger amount of the world's infrastructure today being networked, the risk for a critical cyber warfare attack that could cause disruption in our society is increasing.

During the 2016 US presidential election, cyber attacks against US election infrastructure were reported (Office of the Director of National Intelligence, 2017). These attacks raised concerns about the integrity of that election. Despite these concerns, US officials have been on testimony saying that while the US election infrastructure is possibly not secure, the US election system is too decentralized and distributed for an attacker to change the outcome of an election (Addressing Threats to Election Infrastructure, 2017). However, security expert Halderman (2017) disagree with that statement and claims that attacking the electronic voting machines used around the US would go undetected and hence, lead to an altered outcome of an election.

This paper will investigate, from a cyber warfare approach, the practical conditions needed to manipulate the electronic voting machines to change the outcome of an election.

## 2. Related work

Elections are a central model for collective decision-making – this is why there has been a great amount of research on the subject from a game-theoretic viewpoint. However, these studies mostly focused on the robustness and the computational complexity of manipulating elections with different voting rules. An attempt to manipulate an election is considered resistant for fraud if attacking it is NP-hard. In their paper, Vorobeychik *et al* (2016) studied election fraud against electronic voting machines. Their paper modelled election control as a denial-of-service (DOS) attack on electronic voting machines, with the goal of preventing a certain candidate from winning. They showed that controlling an election by a denial-of-service attack was in P when the protection deployed is deterministic.

## 3. Background

### 3.1 Electronic voting machines

A DRE voting machine displays the ballot for a specific election race on a display in which the voter uses buttons or the display itself (touchscreen) to vote. In order for an authorized voter to vote on a DRE machine, an



This is an excerpt from the Proceedings of the Fourteenth International Conference on Cyber Warfare and Security (ICCWS 2019)

authentication device is needed to be inserted into the machine to start the process. When the authentication device is inserted into the machine, the voter can see the different options for the election on the display. The vote totals are then saved onto the memory card and then summarized in the back-end program associated with the specific machine (Verified Voting, 2017).

The DRE machines could be manipulated by gaining access to the memory cards or the back-end program responsible for the election setup. The authentication devices that are used to unlock the machines for voters could be manipulated or used in other ways to gain privileged access to a machine's sensitive functions.

These devices can also be used for tampering when connected to the back-end program (McDaniel *et al*, 2007).

Several DRE machines have been subject to source code reviews in which all of them were found to have severe flaws and could demonstrate successful attacks (McDaniel *et al*, 2007).

When casting a vote by using an optical scan voting machine, an authorized voter gets a paper ballot. He then enters the marked ballot through the machine, which stores the physical ballot in a ballot box attached to the machine but also records and tallies the vote electronically by the machine's software and stores the votes in its memory components. When the polls close, the machine prints a receipt of the total votes and also saves the votes on a removable media source, e.g. a memory card. The votes on the receipt and the memory card are then compared to verify that the count is correct (Verified Voting, 2017).

The biggest advantage of optical scan machines is that they keep both the physical and the digital records of the votes. In case of an audit, the physical ballots can always be manually recounted. However, since they count votes electronically, the machines are vulnerable to fraud. An attacker can perform the same attack as against a DRE voting machine, but would risk a recount of the paper ballots.

### 3.2 Cyber warfare

The impact of cyber attacks has escalated: they pose a greater danger than ever before. During the last decade, the rise of state-sponsored cyber warfare attacks has caused great concern among security professionals and governments around the world. The anonymous nature of cyberspace makes it easier for nation-states to initiate attacks, since it allows the attacking government to avoid scrutiny from other nations. Cyber warfare is not only limited to nation-states – it only takes a small team of dedicated programmers to find a previously unknown vulnerability to exploit and accomplish substantial damage.

The cost of cyber warfare is hard to determine, and it is subject of debate among experts. Although nobody can know for sure, estimates have been made about different cyber warfare attacks. According to Costin Raiu (2014), the cost of the worm Stuxnet is approximately 100 million dollars and the NetTraveler campaign was estimated to have cost 500 000 dollars.

## 4. Model

The only information available for an attacker is assumed to be the state polls and the type of voting equipment each state used during the 2016 US presidential election. This is publicly available information. The state polls in this paper were gathered from Huffington Post's website (HuffPost Pollster, 2018). The data on voting technology were gathered from Verified Voting's database (Verified Voting, 2017).

To perform an analysis, it was assumed that the state polls accurately reflected voter preferences in that state at the time they were conducted. Further, it was assumed that all conducted attacks were successful, i.e. if an attack were launched against a manufacturer of voting machines, it would always be successful with no defence measures available that could counteract the attack. An attack against an election is associated with a cost that is modelled as directly proportional to the number of manufacturers that the attack targeted.

### 4.1 Poll aggregation

To determine voter preferences of a population, polling organizations take samples that will reflect the demographics of the population of interest. The most accurate poll aggregator during the latest US presidential elections has been Princeton Election Consortium's (PEC) meta-analysis (Wang, 2015).



This is an excerpt from the Proceedings of the Fourteenth International Conference on Cyber Warfare and Security (ICCWS 2019)

The calculation of PEC's meta-analysis is based on state polls which are used to estimate the probability of a Democratic/Republican win. The meta-analysis accomplishes this by using simple statistics mixed with calculating the probability distribution for every possible electoral vote outcome. For all 50 states and the District of Columbia, this amounts to $2^{51} \approx 2.3 * 10^{15}$ different outcomes. The probability of winning is computed by using the median polling margin $M_s$, which is the difference in percentage in support between the two leading candidates, of the three latest state polls. The probability to win $p_s(t)$ for a given state $s$ at time $t$ in the meta-analysis is computed from the polling margin. The estimated standard error of the median (SEM) $\widehat{\sigma_s}$ is calculated to account for the variability of the polls and is defined as,

$$\widehat{\sigma_s} = \left\lfloor \frac{k * MAD}{\sqrt{N}} \right\rfloor \quad (1)$$

where $N$ is the number of polls used, $k$ is a constant scale factor that depends on the distribution. For normally distributed data $k = \frac{1}{\Phi^{-1}(\frac{3}{4})}$ where $\Phi^{-1}$ the inverse of the cumulative distribution function for the standard normal distribution. For a univariate data set $X_i = X_1, \ldots, X_n$, the $MAD$ is the median absolute deviation and is defined as the median of the absolute deviations from the data's median, $MAD = median(|X_i - median(X)|)$. The floor is used to account for intrinsic sampling error and inter-pollster variability and is usually set to 3 if the calculated SEM is smaller than that (Wang, 2015). From the median margin and the SEM a standard score $Z_s = \frac{M_s}{\widehat{\sigma_s}}$ is calculated and translated into win probabilities using the cumulative normal distribution $\Phi(z) = \frac{1}{2}[1 + erf(\frac{z}{\sqrt{2}})]$, where $erf(z)$ is the error function and should be interpreted as the probability that the random variable $Z_i$ falls in the range $[-Z, Z]$.

The probability distribution of all electoral vote outcomes is calculated using the coefficients of $p_s(t)$ from the generating function,

$$P\big(p_s(t)\big) = \prod_s \big(1 - p_s(t)\big) + p_s(t)x^{E_s} \quad (2)$$

where $E_s$ is the number of electoral votes for state $s$. The median of the coefficients from the resulting generating function $P_s(t)$ is used to estimate the number of electoral votes.

#### *4.1.1 Application of meta-analysis*

The basis for the model was the US electoral system to compare how many votes were shifted when manipulating voting systems. An expected outcome of the electoral votes was calculated from the state polls using the meta-analysis. The polls were taken from the start of the year 2016 and the last poll entered were the exit polls at the day of the election. If there were no polls available from the start of the year, instead, the previous election results were used as indication of voter preference until polls for that state were released.

There were some adjustments to the meta-analysis, instead of the median being used, a three-day centered simple moving average was used. This was implemented to pick up trends faster than the median. The difference by doing this compared to PEC's method is when calculating standard error, one must first calculate a three-day centered standard deviation,

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^{3} x_i^2 - 3\bar{x}^2}{3 - 1}} \quad (3)$$

where the 3 arise from the fact that it is a three-point standard deviation. Inserted into the formula for the standard error of the mean yields,

$$\widehat{\sigma_s} = \frac{\sigma_s}{\sqrt{N}} \quad (4)$$

### 4.2 Feasibility study – modelling attack strategy

Suppose that $n$ electoral units $i$ exist in a given country's general election, each corresponding to a number of electoral points $e_i$ (e.g electoral votes or seats in legislative bodies). An attacker would like to change the outcome of the election for the candidate that he wants to win the election by manipulating electronic voting

machines. Throughout, the focus will be on plurality voting, in which a single candidate can be selected by each voter and the candidate with the most votes wins.

Formally, define the election as a pair $E = (C, V)$, where $C = \{c_\alpha, c_1, \ldots, c_m\}$ is a set of candidates, $c_\alpha$ is the attacker's targeted candidate and the subject for election control, and $V = \{v_1, \ldots, v_k\}$ is a set of voters. Every voter has a preference order over the set $C$, where a voter casts a vote for its most preferred candidate. We assume that all voters participate and vote sincerely in the election i.e. there is no strategic voting. Let $\left|V_{c_\alpha}\right|_i$ be the votes for $c_\alpha$ in $i$. We will denote all candidates that is not the attacker's targeted candidate as $c_{-\alpha}$.

**Definition 1.** The winner of the election is the candidate that receives more than half of the total electoral points $\sum_i e_i > \frac{1}{2}\sum e_i$. That is, a candidate needs to win at least $n$ electoral units that has a minimum of $\frac{1}{2}\sum e_i$ electoral points.

**Definition 2.** Enumerate the electronic voting machine models used in electoral unit $i$. The total number of voters in $i$ that uses $l$th electronic voting machine models is denoted $m_{il} = a_{l_i}|V|_{tot_i}$. Here $|V|_{tot_i}$ is the total number of voters in $i$ and $a_{l_i}$ is the fraction of voters in $i$ that uses the $l$th voting machines.

Let $\theta_{ic_m} = \frac{|V_{c_m}|_i}{|V|_{tot_i}}$, denote candidate's strength i.e. to the fraction of voters who rank $c_m$ as their first ranked candidate. This information can in practice be obtained from polls. Let $d_i = \max\left|V_{c_{-\alpha}}\right|_i - \left|V_{c_\alpha}\right|_i$ be the vote difference in $i$ and whenever $d_i < 0$, $c_\alpha$ is expected to win that electoral unit and lose otherwise. If $c_\alpha$ is expected to lose, the attacker decides whether to manipulate or let $c_\alpha$ lose.

Assume that the fraction of voters on $c_\alpha$ that uses $l$ voting machines is equal to the fraction of voters on $c_{-\alpha}$ that uses $l$ voting machines and votes for $c_{-\alpha}$. This allows an attacker to change enough votes on $l$ to change the outcome of the election without raising any suspicion.

**Proposition 1.** *If $d_i < m_{il}\theta_{ic_{-\alpha}}$ it is possible to change the outcome of the election in $i$ by switching votes from one candidate to another by attacking voting machines, and changing the preference of half of the voters on $c_{-\alpha}$ on the $l$th electronic voting model.*

Algorithm 1 shows the election control process by targeting all $i$ that has a vote difference larger than 0. This requirement is enforced under the assumption that an attacker will not try to attack an electoral unit where $c_\alpha$ is expected to win. Proposition 1 is true for all outcomes when $d_i < m_{il}\theta_{-\alpha i}$ and lines 1-4 check whether there exists such an attack where this inequality holds. If no such an attack exists for all candidates $c_{-\alpha}$, election control in $i$ is not possible.

**Algorithm 1** Constructive election control

```
1: Input: d_i, m_il
2: for ∀i do
3:     if d_i > 0 and d_i < m_il θ_ic_m then
4:         return Attack voting systems m_il;
5:     return No manipulation;
```

**Figure 1:** Algorithm 1 displays how an attack against all election units is possible, in which the margin between the two leading candidates is larger than zero

The above algorithm shows a straightforward way of manipulating the election. However, a more risk-averse attacker would not attack everywhere: instead, he would search for election units in which the voter preference is not certain. It is assumed that there are more than one poll available. Then the attacker can obtain the probability that $c_{-\alpha}$ wins $i$. By using PEC's meta-analysis for calculating the probabilities of winning $i$ by using the cumulative normal distribution $\Phi(z_i)$, the attacker can determine which election units can safely be considered to belong to one of the candidates and which ones can be considered uncertain. Uncertain election units have a win probability between 2.5 and 97.5 percent, every $i$ below or above that can be considered to have enough voter support for a candidate to win that unit.



This is an excerpt from the Proceedings of the Fourteenth International Conference on Cyber Warfare and Security (ICCWS 2019)

**Algorithm 2** Constructive election control with risk-aversion

1: Input: $d_i, m_{il}$
2: **for** $\forall i$ **do**
3: $\Phi(z_i) \rightarrow$Calculate win probability
4: **if** $0.025 < \Phi(z_i) < 0.975$ and $d_i < m_{il}\theta_{ic_m}$ **then**
5: **return** Attack voting systems $m_{il}$;
6: **return** No manipulation;

**Figure 2:** Algorithm 2 displays how a risk-averse attacker would only bother to attack uncertain election units in which the probability of winning for candidate $c_{-\alpha}$ is between 2.5 and 97.5 percent

### 4.3 Basic assumptions for staging an attack against electronic voting machines

The cost of manipulating one manufacturer's model of electronic voting machine is based on the EVEREST report (McDaniel *et al*, 2007). In that report, they evaluated vulnerabilities and developed targeted exploits for each model of the voting machines that were subject to that review. To determine the cost of an attack against electronic voting machines, an attacker would need to purchase the equipment required for an attack. This includes electronic voting machines to find vulnerabilities, servers to host virtual machines to test the spread and delivery of the malware, zero-days for operating software and hire competent and reliable personnel. Assuming that the attackers had six months to develop and deploy an attack, an approximation of the cost can be made.

Before any development of exploits for the electronic voting machines can take place, attackers need to decide on a manufacturer or model of voting machines to target. The steps for an attack can also be seen below.

**Purchase of hardware and software** – In order to develop an attack against a specific type of voting machine, an attacker would need to buy the software and hardware associated with that machine. An attacker would also need to purchase servers to run virtual machines on to simulate how the malicious software will spread and make sure it functions properly. To escalate privileges of the malware on the operating software, one or more zero-day exploits is needed. Assuming that the attackers will only focus on the development of malware to voting machines, the attackers purchase zero-day exploits from a vendor.

**Research** – The first team will start with researching the electronic voting machines for vulnerabilities to attack.

**Weaponization** – When the first team has found vulnerabilities to exploit, the second team will develop attacks using these exploits.

**Delivery** – The Delivery team will start at the same time as the second team and will build a payload using the purchased zero-days.

**Test** – A test team is needed to evaluate and report bugs in the developed malware. The testing phase can start when the first package of exploit and payload is bundled together.

**Deployment** – When an exploit and payload have been bundled together and the testing phase has been completed, the malware can be deployed.

## 5. Results

### 5.1 Meta-analysis

Figure 3 shows the expected outcome of electoral votes using the meta-analysis. On the day of the election, the meta-analysis estimated that Hillary Clinton should have received $313 \pm 45$ electoral votes and Donald Trump $225 \pm 45$ votes. The gray bands depict the 95 percent confidence interval and vary in size, since when states with more electoral votes are uncertain, the confidence band will be larger.

The tables below present how many electoral votes (EV) are gained when attacking the different manufacturers and how many different models of each manufacturer had to be targeted. The number of models needed to



This is an excerpt from the Proceedings of the Fourteenth International Conference on Cyber Warfare and Security (ICCWS 2019)

achieve the electoral vote gains when attacking the different manufacturers can be found in the last column. In order to win a US presidential election, a minimum of 270 electoral votes are needed.

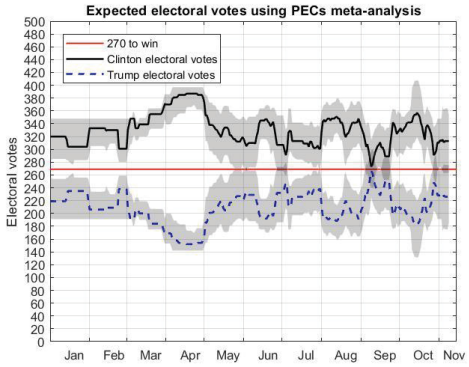


**Figure 3:** A time-series of the expected electoral votes

### *5.1.1 Algorithm 1: DRE voting machines*

The first algorithm had the requirements that the margin between the two leading candidates should be larger than zero and that $d_i < m_{il}\theta_{ic_m}$. As visible in Table 1, the only manufacturers that could have changed the outcome by using Algorithm 1 of the 2016 US presidential election were Danaher, Diebold, Hart, Election Systems & Software and Sequoia.

**Table 1:** Electoral votes shifted when manipulating DRE voting systems

| Brand | Average EV gain | Minimum EV gain | Maximum EV gain | Final EV for Trump | Number of models targeted |
|---|---|---|---|---|---|
| Avante | 0 | 0 | 0 | 225 (0) | N/A |
| Danaher | 21.66 | 3 | 23 | 248 (23) | 1 |
| DFM | 0 | 0 | 0 | 225 (0) | N/A |
| Diebold | 27.09 | 0 | 60 | 245 (20) | 1 |
| Dominion | 0 | 0 | 0 | 225 (0) | N/A |
| ESS | 36.62 | 20 | 67 | 260 (35) | 1 |
| Hart | 5.43 | 0 | 38 | 225 (0) | 1 |
| MicroVote | 0 | 0 | 0 | 225 (0) | N/A |
| Sequoia | 31.32 | 14 | 51 | 265 (40) | 2 |
| Unilect | 0 | 0 | 0 | 225 (0) | N/A |
| Unisyn | 0 | 0 | 0 | 225 (0) | N/A |

From the above results it is obvious that an attacker would have the highest rate of success if he would target Diebold, Election System & Software and Sequoia. However, none of these change the final outcome of the election. If the attacker would launch an attack against all these manufacturers at once, Donald Trump would have gotten 296 electoral votes as a final result and would have won the election. This would mean that the attackers would have had to target four different models of electronic voting machines. As visible in Figure 2, attacking DRE voting machines could only have changed the outcome of the election at certain points during the year 2016.

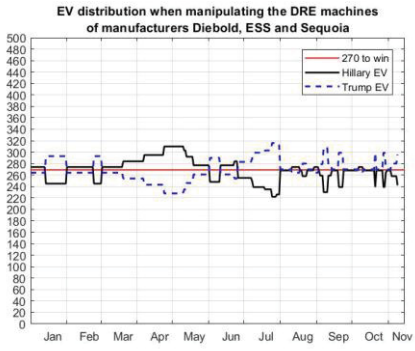


**Figure 4:** A time-series of the electoral votes when an attack is carried out against Diebold, Election System & Software and Sequoia

### *5.1.2 Algorithm 1: Optical scan voting machines*

Table 2 displays how an attacker could manipulate the optical scan voting machines of Election System & Software and gain a minimum of 96 electoral votes. To make this happen, an attacker would need to target three models of optical scan voting machines from Election Systems & Software. Although, if they did that, they would always have been able to control the election. Figure 4 displays how electoral votes would have been distributed during the 2016 US presidential election if an attack had been carried out against Election System & Software.

**Table 2:** Electoral votes shifted when manipulating optical scan voting machines.

| Brand | Average EV gain | Minimum EV gain | Maximum EV gain | Final EV for Trump | Number of models targeted |
|---|---|---|---|---|---|
| Avante | 0 | 0 | 0 | 225 (0) | N/A |
| Danaher | 0 | 0 | 0 | 225 (0) | N/A |
| DFM | 0 | 0 | 0 | 225 (0) | N/A |
| Diebold | 22.3 | 4 | 65 | 258 (33) | 2 |
| Dominion | 16.18 | 5 | 57 | 259 (34) | 1 |
| ESS | 145.94 | 96 | 185 | 339 (114) | 3 |
| Hart | 1.89 | 0 | 31 | 225 (0) | 1 |
| MicroVote | 0 | 0 | 0 | 225 (0) | N/A |
| Sequoia | 17.43 | 0 | 55 | 264 (39) | 2 |
| Unilect | 0 | 0 | 0 | 225 (0) | N/A |
| Unisyn | 12.58 | 0 | 29 | 225 (0) | 1 |

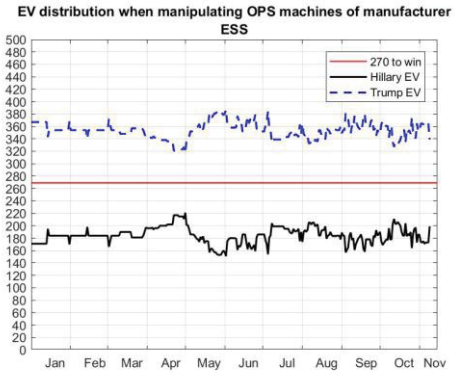


**Figure 5:** Electoral vote distribution when attacking three models of optical scan voting machines from Election Systems & Software. As visible in the figure, Donald Trump would have won the election during any time of the campaign

Comparing Figure 3 and Figure 5, it is noticeable that attacking the optical scan machines from Election Systems & Software would have had a great impact on the election.

### *5.1.3 Algorithm 2: DRE voting machines*

The evaluation of the second algorithm where the attacker only attacked when the win probability for $c_{-\alpha}$ is between 2.5 and 97.5 percent can be seen in Table 3.

**Table 3:** Electoral votes shifted when manipulating DRE voting systems.

| Brand | Average EV gain | Minimum EV gain | Maximum EV gain | Final EV for Trump | Number of models targeted |
|---|---|---|---|---|---|
| Avante | 0 | 0 | 0 | 225 (0) | N/A |
| Danaher | 11.72 | 0 | 20 | 245 (20) | 1 |
| DFM | 0 | 0 | 0 | 225 (0) | N/A |
| Diebold | 21.98 | 0 | 44 | 245 (20) | 1 |
| Dominion | 0 | 0 | 0 | 225 (0) | N/A |
| ESS | 24.68 | 0 | 67 | 260 (35) | 1 |
| Hart | 5.38 | 0 | 38 | 225 (0) | 1 |
| MicroVote | 0 | 0 | 0 | 225 (0) | N/A |
| Sequoia | 13.99 | 0 | 37 | 245 (20) | 2 |
| Unilect | 0 | 0 | 0 | 225 (0) | N/A |
| Unisyn | 0 | 0 | 0 | 225 (0) | N/A |



This is an excerpt from the Proceedings of the Fourteenth International Conference on Cyber Warfare and Security (ICCWS 2019)

Comparing the two algorithms, it is noticeable that the maximum electoral vote gain is mostly the same. However, the average and minimum electoral vote gain is much smaller, which is expected since the second algorithm only attack states that are considered uncertain.

### *5.1.4 Algorithm 2: Optical scan voting machines*

Table 4 presents the electoral votes that have been shifted when attacking optical scan machines with Algorithm 2. Compared with Algorithm 1, the average gain over the campaign has gone significantly down. Although, even when only attacking uncertain states, Election Systems & Software could have changed the outcome of the election.

**Table 4:** Electoral votes shifted when manipulating optical scan voting machines.

| Brand | Average EV gain | Minimum EV gain | Maximum EV gain | Final EV for Trump | Number of models targeted |
|---|---|---|---|---|---|
| Avante | 0 | 0 | 0 | 225 (0) | N/A |
| Danaher | 11.72 | 0 | 20 | 245 (20) | 1 |
| DFM | 0 | 0 | 0 | 225 (0) | N/A |
| Diebold | 21.98 | 0 | 44 | 245 (20) | 1 |
| Dominion | 0 | 0 | 0 | 225 (0) | N/A |
| ESS | 24.68 | 0 | 67 | 260 (35) | 1 |
| Hart | 5.38 | 0 | 38 | 225 (0) | 1 |
| MicroVote | 0 | 0 | 0 | 225 (0) | N/A |
| Sequoia | 13.99 | 0 | 37 | 245 (20) | 2 |
| Unilect | 0 | 0 | 0 | 225 (0) | N/A |
| Unisyn | 0 | 0 | 0 | 225 (0) | N/A |

Attacking the optical scan systems of Election System & Software could have changed the outcome at several points during the year 2016. But not at all points as with Algorithm 1, which can be seen in Figure 6.

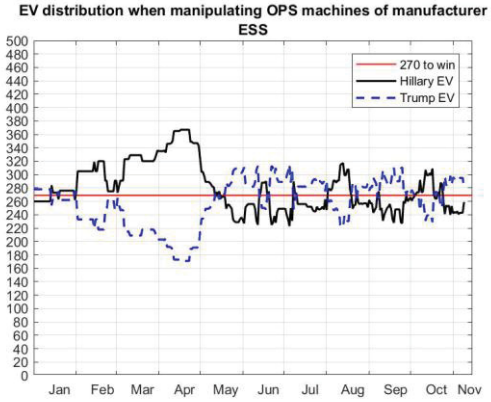


**Figure 6:** Electoral vote distribution when attacking three models of optical scan voting machines from Election Systems & Software

## 5.2 Cost of staging an attack against electronic voting machines

Table 5 presents the different costs involved in the different steps for attacking a specific model of electronic voting machines of one manufacturer. The purchase of software and hardware, items 1-3 in Table 1, includes the purchase of servers to host virtual machines for testing the malware on several machines. By using VMware's cost calculator (VMware, 2018) for two servers that can host two hundred virtual machines, the total amount is USD 300 000. The cost of the voting machine hardware and software is taken from Elections Systems & Software order form (State of Kansas, 2018) and is based on the purchase of ten pieces of the same model of optical scan machines with the associated software. The last cost for purchase of hardware and software is three zero-days for the Windows operating software, purchased from Zerodium (Zerodium, 2018). The most substantial item in Table 5 is personnel costs. The Research team is based on the EVEREST report (McDaniel *et al*, 2007). The rest of the personnel costs are based on our own estimates including infrastructure costs. The assumption was that this would be a six-month operation. The total number of individuals is assumed to be approximately sixty, with most of them working on testing. The unit cost is based on two hundred dollars per hour multiplied by a month's

work (160 hours). The total sum to manipulate three models of electronic voting machine from one manufacturer amounts to 10 354 000 dollars.

**Table 5:** The table presents the different costs of developing a cyber warfare attack against electronic voting machines

| Description | Sum of post |
|---|---|
| Virtual machines | $ 300 000,00 |
| Voting machines hardware and software | $ 200 000,00 |
| Zero-days | $ 990 000,00 |
| Research | $ 1 344 000,00 |
| Weaponization | $ 3 840 000,00 |
| Delivery | $ 1 920 000,00 |
| Test | $ 1 600 000,00 |
| Deployment | $ 160 000,00 |
| **Total sum** | $ 10 354 000,00 |

## 6. Conclusion and discussion

This paper has investigated, from an attacker's perspective, the effort needed to manipulate the electronic voting machine to change the outcome of a US presidential election. The goal was to determine whether it was possible for an attacker to hack an election with regards to the risk and cost.

This paper demonstrated how an attacker might build a strategy to attack electronic voting machines using two realistic scenarios. Both scenarios have the same basic criteria and assumptions: in order to be able to win a state, the number of people that do not vote for the attacker's targeted candidate and use electronic voting machines must be greater than the margin of votes. The underlying assumptions for both scenarios were that the state polls accurately reflected voter preferences at the time they were conducted and that all attacks were successful.

The model shows that it was feasible during the 2016 US presidential election to change the outcome of that election by hacking the electronic voting machines. For example, targeting three models of optical scan machines manufactured by Election Systems & Software could have made it possible to gain at least 96 electoral votes. Since such an attack could always have controlled the election during the year, an attacker could control the electoral vote outcome by using the model presented. An attacker could use the presented model in order to see how many electoral votes had to be gained starting from when the primary candidate for each party has been elected. The attack would have been possible at a cost of ten million dollars. Since this cost is low compared to what it achieves, it becomes not only a venture for nation-states but could be used by private actors as well.

The major finding of this article is that attacking models by one manufacturer of DRE voting machines alone, could not have changed the election. Attacking models of optical scan voting machines however, would always have made it possible by targeting three models from Election System & Software. Attacking optical scan machines is a greater risk for an attacker and is not be the best strategy. One must add that using the first algorithm is not a realistic approach and would most likely be detected. The second algorithm, which only targets uncertain states, is a more realistic approach and could more likely go undetected.

It is important to notice that the article is not based on the assumption that the 2016 US presidential election was hacked. The article seeks to introduce a new way of thinking how easy and cheap it can be to hack an election. The model only considers the scale of an attack and not specific targets. This kind of model can be used to measure other kinds of targeted cyber warfare attacks against critical infrastructure to estimate the costs and effect. The conclusion that is drawn is that that election was most likely not hacked, but the option cannot be excluded. The only recounts that were completed were cases when the ballots were recounted by re-entering them through optical scan machines, which – if they were manipulated – would not give a different outcome. All elections that use digital solutions have this imminent threat since all software, no matter how secure, have vulnerabilities that can be exploited. There is not a way for a digital solution to be secure enough to be used in an election. With electronic voting machines an attack vector is introduced that makes the use of these machines not acceptable in elections. Paper ballots should be used instead, as they are more secure and reliable.

### 6.1 Further work

To further improve the results of the model, an algorithm could be implemented that minimizes the attack target by attacking a minimal number of electronic voting machines to give the largest possible electoral votes. If county level electronic machine use and voting history would be implemented, the results could be more precise, and the attack target could be minimized. Another approach that could minimize the attack target is to implement a disinformation strategy that could change voter preferences in volatile states. This would make the states less volatile and minimize the number of machines needed to be attacked.

Another interesting approach would be to implement a Stackelberg security game, similar to the one described in Vorobeychik et al (2016) paper. This would introduce a randomized defence for election officials that could be modelled as random audits. An attacker would then have to randomize his strategy to circumvent the audits in order to stay undetected.

This paper is based on poling data as a publicly available source of information aimed at giving an indication on the expected outcome of an election. It would be interesting to construct and evaluate similar models based on other types of input data.

## References

Addressing Threats to Election Infrastructure: Hearings before the Select Committee on Intelligence, US Senate, 115th Congress. (2017) (Testimony of Jeanette Manfra and Dr. Samuel Liles).

HuffPost Pollster, 2018, HuffPost Pollster, viewed 30 August 2018, http://elections.huffingtonpost.com/pollster.

Kansas Secretary of State. (2005). Election Systems & Software Order Form. Retrieved from https://www.kssos.org/other/clerks/Voting_Equipment/ES&S%20order%20form.xls.

McDaniel et al. (2007) EVEREST: Evaluation and Validation of Election-Related Equipment, Standards and Testing: report to the Ohio Secretary of State.

Office of the Director of National Intelligence. (2017) Assessing Russian Activities and Intentions in Recent US Elections.

Russian Interference in the 2016 US Election: Hearings before the Select Committee on Intelligence, US Senate, 115th Congress. (2017) (Testimony of J. Alex Halderman).

Verified Voting, 2017, Verified Voting, viewed 30 August 2018, https://www.verifiedvoting.org/verifier/#year/2016/.

VMware Inc, 2009, VMware Inc, viewed 14 September 2018, https://www.vmware.com/support/.../doc/vCenter_Chargeback-Costing_Calculator.xls.

Wang S. S-H. (2015), 'Origins of Presidential poll aggregation: A perspective from 2004 to 2012', International Journal of Forecasting, vol. 31, Issue 3, pp. 898-909.

Zerodium, 2018, Zerodium, viewed 14 September 2018, https://zerodium.com/program.html.