Sweeney L, Yoo J, Zang J. Voter Identity Theft: Submitting Changes to Voter Registrations Online to Disrupt Elections. *Technology Science*. 2017090601. Version 2. September 6, 2017. http://techscience.org/a/2017090601

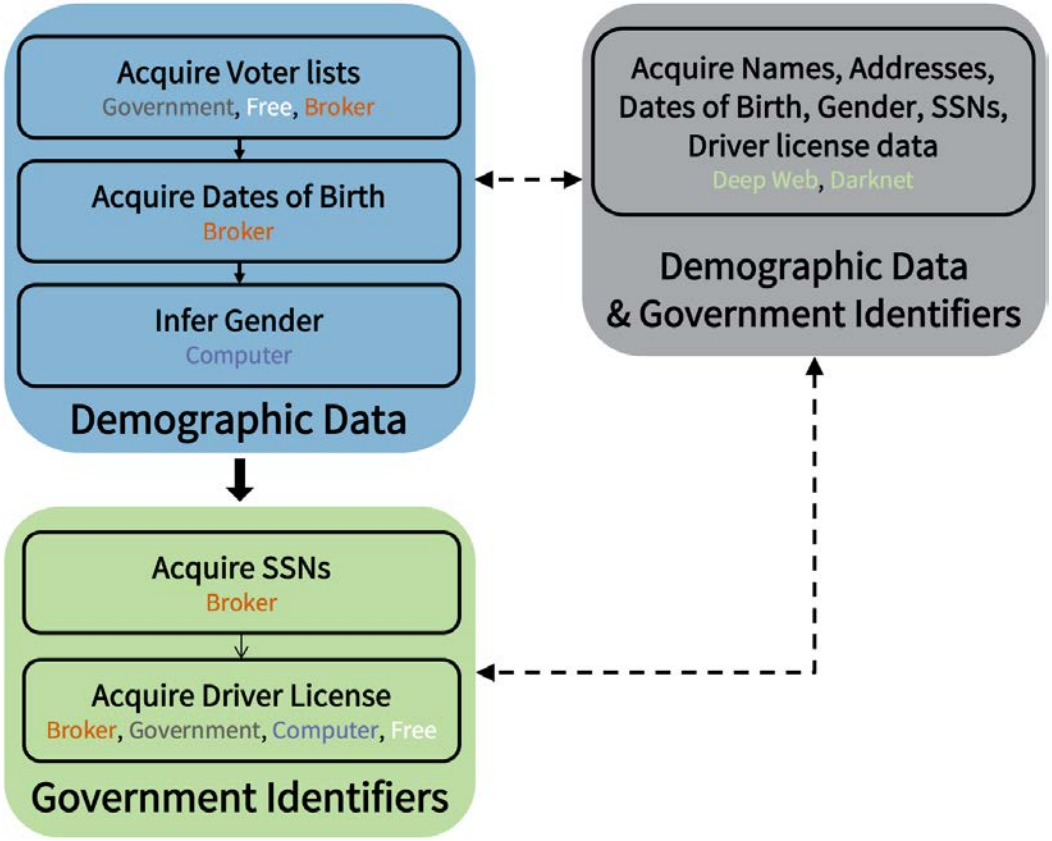


Voter Identity Theft: Submitting Changes to Voter Registrations Online to Disrupt Elections

Latanya Sweeney, Ji Su Yoo, and Jinyan Zang

Highlights

- Websites for 35 states and DC in 2016 were vulnerable to voter identity theft attacks: an imposter could submit changes to voter registration information
- An imposter needed a combination of voter's name, date of birth, gender, address, Social Security Number, or Driver's License Number
- Relevant data can be acquired from government, data brokers, or darknet markets. Total cost of an automated attack against 1% of all vulnerable voter registrations nationwide ranged from $10,081 to $24,926 depending on the data source used. States cost less, e.g., $1 for Alaska and $1,020 for Illinois
- A voter identity theft attack could disrupt an election by imposters submitting address changes, deleting voter registrations, or requesting absentee ballots



Ways an attacker can acquire the personal data needed to impersonate voters.

Abstract

Could an attacker impact U.S. elections by merely changing voter registrations online? This reportedly happened during the 2016 Republican primary election in Riverside County, California. What about elsewhere? We surveyed official voter record websites for the 50 states and the District of Columbia and assessed the means and costs for an attacker to change voter addresses. Relatedly, an attacker could also change party affiliations, delete voter registrations, or request absentee ballots online. A voter whose address was changed without her knowledge, for example, in most states would have a polling place different than expected. On Election Day, when she appeared at her presumed polling place, she would have been unable to cast a regular vote because her name was not on the precinct's register. She may have been turned away or given a provisional ballot, and in many cases, a provisional ballot would not count. Perpetrated at scale, changing voter addresses, deleting voter registrations, or requesting absentee ballots could disenfranchise a significant percentage of voters, and if carefully distributed, such an attack might go unnoticed even if the impact was significant. So, how practical is it to submit false changes to voter registrations online?

Results summary: We found that in 2016, the District of Columbia and 35 of the 50 states had websites that allowed voters to submit registration changes. These websites determined whether a visitor was an actual voter by requesting commonly available personal information. Some websites gave multiple ways for a voter to self-identify. Of these, {name, date of birth, address} was required in 15, {name, date of birth, driver's license number} was required in 27, and {name, date of birth, last 4 SSN} was required in 3. We found that an attacker could acquire the voter names, demographic information and government-issued numbers needed to impersonate voters on all 36 websites from government offices, data brokers, the deep web, or darknet markets.

Overall, the total cost of an attack varied based on the number of voters to impersonate, data sources used, whether the websites had CAPTCHAs, and specific states of interest. We found that the practical costs of changing 1 percent of the voters on all 36 websites could range from $10,081 to $24,926 depending on whether the attacker used data from government, data broker, darknet or other sources. Costs for an attack on a specific geographical area or state were much less, such as $1 for Alaska or $1,020 for Illinois. Back office processes and election practices, which varied among states, could have possibly limited attack success rates.

Keywords: election, voting, identity theft, authentication, computer security, privacy

Introduction

Imagine impersonating a voter online in order to submit voter record changes to prevent the voter from voting. Now imagine doing that for hundreds or thousands of voters in an attempt to impact outcomes or to add uncertainty or chaos to the voting franchise. The idea may sound far-fetched, but local area newspapers reported that one such incident occurred in Riverside County, California in the United States before the Republican Party primary on June 7, 2016 [1], [2]. The Riverside County District Attorney confirmed that someone with access to voters' Social Security numbers and other personal information went to the state's voter registration website, pretended to be a voter, changed that voter's party affiliation, then did so seemingly for up to hundreds of voters without the knowledge of those voters [1].

Changing the party affiliation of a voter before the Republican primary on June 7 was especially problematic because only registered Republicans can vote in the Republican primary in California, and reportedly some voters only learned about their registration change at the polling place. Most of those voters received provisional ballots rather than regular Republican ballots at the polls, but those provisional ballots did not count because the voters were not registered as Republicans [1]. Some voters saw their affiliation go from Republican to the Green Party [2]. The county branch of the Republican Party called attention to the issue after receiving a couple hundred calls from individuals who had their party affiliation changed with no notification process [2]. Only Republican Party registrations seemed affected. Investigations were made more difficult by the fact that California state

officials reportedly did not retain sufficient website and database logs that would identify the Internet addresses of the machines visiting the California voter registration website [1], [2].

This was one primary whose investigation remains open based on a factual basis for allegations. Regardless of the outcome, it begs the question: could hundreds of thousands of voters around the United States have experienced the same kind of disruption at the polling place on November 8, 2016 during the general presidential election in the United States that Republican voters in Riverside County reportedly experienced on June 7, 2016, during the Republican primary? The answer depends on how many state websites allowed voters to submit changes to their own voting records online, whether those websites were vulnerable to attack, and if attacked, whether back office processes or state practices could have possibly limited the attack.

In this writing, we make a sharp distinction between registering new voters and changing existing voter registrations online. Voters can change their records online even if they did not originally register online. The ability to edit voter records is offered by many voter registration websites and some driver's license websites, depending on the state. So, this paper does not find a concern specific to voter registration websites, only with the ability to submit changes to voter records online, regardless of the purpose of the website.

This paper assesses what is required for an attacker to impersonate a voter at a state website and then submit a change to the voter's personal information that would either render the voter unable to vote on Election Day or keep the voter's ballot from being fully counted. This paper ends with a Discussion section, which summarizes significant results, broader impacts, and examples of states' current back office practices. A reader could advance to the Discussion section, at the end of this paper, for an expedited read and return to the body of the paper for details as needed.

Elections rely on trust and transparency. Technological vulnerabilities will emerge and when they do, it is important that they be addressed swiftly and understood publicly in order to maintain public faith in the democratic process. Ignored and secret problems only enable attackers to exploit vulnerabilities freely.

Background

Online Voter Registration

In many states, the ability to change existing voter registrations online is provided along with the ability for new voters to register online ("online voter registration" in this writing). The first state with online voter registration was Arizona in 2002, followed by Washington six years later [3]. There were 11 more states offering online registration by 2012, 7 more by July 2014 [3], and a reported total of 31 states and D.C. by June 2016 [4]. Kentucky was among the most recent, launching its system on March 1, 2016 [5].

One of the major motivations for states to provide an online voter registration tool is to reduce cost. While an online tool costs approximately $240,000 on average to build, with the highest cost of $1.8 million for California [4], [6], voter registration websites can significantly reduce costs. For example, Arizona reported a per-registration processing cost of 83 cents for paper registrations but only 3 cents for online registrations [4]. Beyond cost reductions, improved accessibility and visibility of registration data are additional benefits of online registration systems to voters and election administrators [7].

In January 2014, the Presidential Commission on Election Administration (PCEA) recommended that U.S. "states adopt online voter registration" [8] in order to reduce errors from paper-based registration, save money, decrease delays at the polling place with more accurate rolls, and improve the voter registration experience.

In a 2014 survey of 8,000 local election officials, researchers found that moving to online registration was one of the most common responses to an open-ended question about how to improve the election administration process. Similar comments were made repeatedly in public hearings of the PCEA [9].

While encouraging states to deploy online voter registration systems, the PCEA report also cautioned that "questions about security will require close attention to ensure that unauthorized changes to voter registration cannot be made" [8]. These concerns were reiterated by the Congressional Research Service (CRS) in a report on October 18, 2016, which stated that "successful attacks could compromise the confidentiality, integrity, or availability of election information or processes…For example, voter registration lists could be deleted or altered" [10]. If so, how could it be done?

An attacker who pretends to be a voter online engages in a form of identity theft. The attacker's goal is to disenfranchise the impersonated voter or to discredit the election system, which is possible if the impostor can convince the website that he is the actual voter and is therefore eligible to submit online changes to the voter's registration record and his submitted changes are accepted. How does the website know a visitor is a particular voter? A website considers anyone, or anything, that provides the correct personal information about a voter to be the voter. So, an impostor needs to have personal information about voters in order to impersonate voters online.

The paragraphs below describe identity theft, sources of personal information, ways that changing voter registrations can disenfranchise voters or undermine confidence in elections, and the kinds of back office logs state administrators may keep to limit an attacker or help track one down. Afterwards, we use these concepts to figure out what sources of personal information and which websites would have allowed impostors to submit changes to voter registrations in the 2016 presidential election.

Voter Identity Theft

Identity theft is the fraudulent use of a person's identifying information, usually for financial gain. We introduce the term "voter identity theft" (or "voter ID theft") to describe an impostor's use of a voter's personally identifying information to impersonate the voter in the voting franchise. The Riverside County, California case appears to be the first reported case of a cyber-attack using voter ID theft [1], [2].

Other government services encounter identity theft as a challenge as they operate and secure systems to serve the needs of Americans. For example, the Internal Revenue Service (IRS) regularly confronts identity theft in connection with tax fraud. The attacker impersonates a taxpayer and files a false tax return to obtain a tax refund before the actual taxpayer files her own legitimate return. When the actual taxpayer files subsequently, she receives a notice of double filing from the IRS and begins a lengthy administrative review process [11]. According to the U.S. Government Accountability Office in a report to Congress, the IRS estimates that in 2014, it paid $3.1 billion in refunds for 1.3 million false tax returns filed by identity thieves [11]. Anti-identity theft measures such as the IRS's Identity Protection PIN and e-File PIN, which were given only to taxpayers who have already been victims of identity theft, were found to have been compromised by identity thieves and were suspended in 2016 [12], [13]. In May 2015, the IRS suspended its Get Transcript service, which allows taxpayers to view their old tax returns online, after identity thieves attacked the service and used the personal information of 100,000 taxpayers to log in and acquire the taxpayers' returns [14].

Cyber-attacks involving identity theft do not require technical penetration of the computers that power the websites, nor do they require the kind of computer "break-ins," data breaches, or compromised passwords traditionally discussed as computer security concerns. Instead, cyber-attacks involving identity theft rely on access to personal information, which today is widely available on Americans. Impersonating a registered voter online merely requires having personal data on the voter and knowledge of election specifics. It involves little or no computer hacking expertise.

Attacks based on Voter ID

Without breaking into a computer or compromising a password, voter ID theft can disrupt voting by changing voter registrations. Here are three ways.

1. The attacker can change a home address to make a voter ineligible to vote in the local precinct (or county) where the voter expected to vote in most states [15]. We term this the "change of address attack."

2. The attacker can request an absentee ballot, making a voter ineligible to vote in person such as in some counties in Florida [16], or unless the voter remembers to bring the blank absentee ballot to the polling place in states such as in California or Virginia [17], [18]. We term this the "absentee ballot attack."

3. The attacker can change party affiliation to a different party to make a voter ineligible to vote in closed party primaries as reported in the Riverside County incident described earlier.

An attack can combine the first two methods by first changing a voter's address to an attacker-preferred address and then requesting that an absentee ballot be sent to the new address. In this case, the attacker might submit a false absentee ballot on behalf of the voter. This was one of the main concerns of security researchers regarding Maryland's decision on September 14, 2016 to move to a new online ballot system [19], [20], [21].

Changes to voter records made by an attacker can have long-term consequences in many states. According to the National Voter Registration Act (NVRA) of 1993, a state is allowed to remove voters from local registration rolls due to address change, mental incapacity, criminal conviction, or death [22], with address change generally the only option that publicly accessible state online websites offer to voters.

Provisional ballots

A malicious attack that changes a voter's home address in most states can adversely impact the voter's experience at the polling place. The voter may be prohibited from voting, required to somehow locate and visit the precinct assigned to his new address—an address that is unknown to him—or obliged to cast a provisional ballot at the precinct that is no longer assigned to him [23]. If the new address set by the attacker is still in the same Congressional district as the old address of the voter, then the NVRA requires states to permit the voter to vote either at the previous polling place, the new polling place, or a designated central location upon "oral or written affirmation by the registrant of the new address." A voter will find it difficult to affirm a new address if the new address on file results from an attack rather than an actual move [22]. If the attacker chooses a new address in a different Congressional district, then the voter would likely receive a provisional ballot at her old precinct on Election Day, but provisional ballots are often not counted, depending on the state and circumstances.

The National Conference of State Legislators (NCSL) in the United States conducted a survey in June 2015 of the nation and found 24 states (updated to exclude Illinois) do not count provisional ballots that were cast in the wrong precinct [23]. Illinois counts provisional ballots in some cases. Table 1 has a summary. Twenty-one states and D.C. include in their vote tallies a partial count of provisional ballots cast in the wrong precinct by only counting votes for federal races, statewide races, or local races that are shared between the new and old precincts [23]. Maine does a full count of all races from all provisional ballots cast in the wrong precinct only if the number of provisional ballots cast is large enough to affect the outcome of an election [23]. Idaho, Minnesota, and New Hampshire do not issue provisional ballots since they offer same-day registration. North Dakota does not have voter registration

and only issues provisional ballots when poll hours are extended [23]. In 2016, 2.5 million provisional ballots were cast nationally, and 62 percent were counted in full [24].

Handling of Provisional Ballots Cast in the Wrong Precinct	State
Full Count	Maine
Partial Count (Only count federal races, state-wide races, or shared races between precincts)	Alaska, Arizona, Arkansas, California, Colorado, District of Columbia, Georgia, Illinois, Kansas, Louisiana, Maryland, Massachusetts, New Jersey, New Mexico, New York, Ohio, Oregon, Pennsylvania, Rhode Island, Utah, Washington, West Virginia
Do Not Count	Alabama, Connecticut, Delaware, Florida, Hawaii, Indiana, Iowa, Kentucky, Michigan, Mississippi, Missouri, Montana, Nebraska, Nevada, North Carolina, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Vermont, Virginia, Wisconsin, Wyoming

Table 1. State Handling of Provisional Ballots Cast in the Wrong Precinct. From the National Conference of State Legislators in June 2015 [23] corrected by Illinois election official. Idaho, Minnesota, New Hampshire, and North Dakota are excluded for having either same-day registration or no registration.

Timing matters

For a hypothetical change of address attack to succeed, the attacker would need to focus on those states that host websites that allow changes to voter rolls and perpetrate the attack on those websites before registration deadlines. According to the NVRA, states may set registration deadlines of up to 30 days before an election [22]. In the days before the deadline, voters can submit personal information to the state on physical or electronic voter forms for change of address, which the state can then act on by immediately updating the voter rolls before the election [22]. After the registration deadline, states close the voter rolls to updates until after the election. Voters are expected to vote at the designated precincts based on the addresses that appear on the rolls [25].

In order for a falsified absentee ballot attack to succeed, an attacker needs to request the ballots on behalf of voters before the absentee ballot request deadline of each state. After the deadline passes, voters filing absentee ballot requests may be required to vote in person. Deadlines for absentee ballot requests range from 21 days before the election for Rhode Island to noon of the day before for 7 states [26].

In the case of the Riverside County attack, described earlier, registration changes may have occurred as early as two months before the election. At least one voter reported seeing his party affiliation changed online when he checked on April 11, almost two months before the

June 7 primary [1]. However, complaints did not become prominent until the Republican primary on June 7, reportedly when voters found themselves receiving unexpected provisional ballots at the polling place.

Authenticating voters

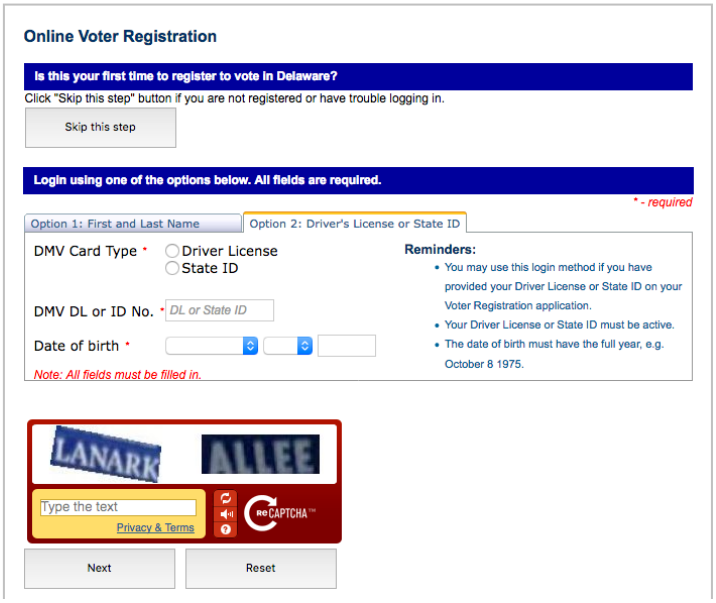
A critical component to the change of address attack and the absentee ballot attack is the personally identifying information a state voter website requests from a visitor to prove he is the actual voter.

A May 2015 survey by Pew of 20 states found that each of the states surveyed used a combination of either driver's license Number/State ID number, last 4 digits of the Social Security Number (SSN), and/or the full Social Security Number to authenticate website visitors for voter registration [3]. Louisiana also required the audit code of the driver's license or State ID in addition to the driver's license Number or State ID number, and Washington required the issuance date of the driver's license or State ID in addition to the driver's license Number or State ID number [3].

Figure 1 shows the voter registration pages for Delaware at which a Delaware voter can change home address. In Delaware, a registered voter may enter either name, date of birth, and 5-digit ZIP (postal) code, as shown in Figure 1a, or driver's license number (or State ID) and date of birth, as shown in Figure 1b.



(a)



(b)

Figure 1. Information required to identify a voter at the Delaware website is either: (a) name, date of birth, and ZIP; or, (b) driver's license number (or State ID if not a driver) and date of birth. In the example shown, the voter also has to enter "LANARK" in the CAPTCHA field to proceed. Visiting the web page again would require the voter to enter different text in the CAPTCHA field. See Figure 4 for the source URL.

Sources of Personal Information on Americans

The intent is to require voters to provide "information that others will not have," according to the bipartisan National Conference of State Legislators [4]. However, in today's data-rich society, is it reasonable to believe that a Social Security number or driver's license number, or in the Delaware example (Figure 1) date of birth, is only known to the voter?

Social Security numbers (SSNs) are issued by the federal U.S. Social Security Administration to people who work in the United States, babies born in the United States, and people who are tax dependents of U.S. workers. Driver's license numbers refer to the numbers that appear on identification cards issued by states as a requirement to drive a motor vehicle in the United States. Many states issue identification cards, through the same department that issues driver's license numbers, to those state residents who do not drive.

Below are four readily available ways of procuring personally identifying information on Americans, including SSNs and driver's license numbers.

1. Purchase from a data broker or vendor

There is a significant ecosystem of data brokers and vendors that legally acquire sensitive personal information such as Social Security Numbers or driver's license numbers to sell to clients for background checks, fraud checks, or other purposes [27]. These companies include credit report bureaus such as Experian and data brokers such as Acxiom [28], [29]. Some companies sell their data to others who may then, in turn, use it for identity theft purposes [30],[31]. For example, in 2013, Experian discovered that one of its subsidiaries, Court Ventures, sold 3.1 million personal records of Americans including Social Security numbers, dates of birth, addresses, previous addresses, phone numbers, email addresses and other sensitive data to a Vietnam-based identity theft service, Superget.info, from at least 2011 to February 2013 [30]. The Vietnam company earned $1.9 million by selling its subscription to Court Venture's data to its 1,300 clients for use for identity theft [30]. In February 2016, the Federal Trade Commission won $5.7 million judgments against two data brokers, LeapLab and Leads Company, for gathering personal information from payday lending applications. The information gathered was a consumer's name, address, phone number, employer, Social Security number, and bank account number, including the bank routing number. These companies sold the data to non-lenders, who had no need for this information, for $0.50 a record [30].

2. Acquire data from a data breach, directly or indirectly

In recent years, multiple major data breaches have affected millions of Americans. A 2015 survey by the American Institute of CPAs found that 25 percent of Americans fell victim to information security breaches in the past year [32]. The number of data breaches, as tracked by the Identity Theft Resource Center, reached an all-time high in 2015 of 338 breaches involving Social Security Numbers and a combined total of 164.4 million records [33]. Victims of major breaches include: (a) the 21.5 million current and former federal workers affected by the Office of Personnel Management breach in July 2015 [34]; (b) the 78.8 million current and former customers and employees of Anthem, a health insurance company, who had their names, birth dates, Social Security numbers (SSNs), home addresses, and other personal information stolen in an attack on February 2015 [35]; and, (c) the 15 million customers of T-Mobile, the wireless service provider, who had their names, dates of birth, addresses, Social Security numbers, and driver's license numbers exposed through a breached Experian server that T-Mobile used for credit assessment in October 2015 [36]. Even public figures such as former First Lady Michelle Obama, former Vice President Al Gore, and the singer Beyoncé have had their credit reports and Social Security numbers posted online in March 2013 as a result of attacks against the three credit bureaus Experian, Equifax, and TransUnion [37].

These breaches and many documented others provided personal data on Americans sufficient to enable black-market vendors to sell personal credentials such as Social Security numbers at fairly low prices. One survey in August 2016 found Social Security numbers sold discreetly on the web for $1 per record [38]. Security researchers also found web sites such as SSNDOB that sell Social Security numbers for $1 and driver's license

number for $4 [39]. In two years, SSNDOB served more than 1.02 million Social Security numbers and more than 3.1 million birth records to its customers [40]. SSNDOB allegedly acquired the personal data by infiltrating the internal systems of data brokers LexisNexis and Dun & Bradstreet and the background check company HireRight [40].

There have been breaches of voter data from state agencies and from political campaigns. In June 2016, the United States Federal Bureau of Investigation (FBI) investigated a hacking attack against a database of 15 million voter records at the Illinois Board of Elections, with the attackers using a SQL injection to steal tens of thousands of records [41], [42], [43]. The database was 10 years old and included voter's names, addresses, sex, birthdates, with some records including the last 4 digits of the voter's SSN or the voter's entire driver's license number. Board of Elections officials declared that they were "highly confident [the hackers] weren't able to change anything" [42]. Investigators believe the hackers were likely based overseas, and suggested Russia as a possibility [44].

In May 2016, Arizona officials took the statewide voting registration system offline after the FBI alerted the Arizona Department of Administration of a credible threat [41]. Arizona officials discovered that a county election official's username and password had been posted online. That official's account could only access county-level data rather than the entire state system. State investigations of the incident found no evidence that any data in the system was compromised or any malware installed.

Georgia had two significant breaches of voter information in about the same number of years. A breach of about 7.5 million Georgia voter records took place in March 2017 [45]. In 2015, the Office of the Georgia Secretary of State accidentally sent the Social Security numbers and other private information of more than 6 million Georgia voters to a distribution list of media outlets and political parties [45].

3. Predict driver's license numbers from other personal information

Intrinsic vulnerabilities in the design of some driver's license numbers may allow others to predict an individual's driver's license number from publicly available personal information such as the individual's name and date of birth.

University of Wisconsin–Madison researcher Alan De Smet has published online the methodology to predict the driver's license numbers of 11 states: Florida, Illinois, Maryland, Michigan, Minnesota, Nevada, New Hampshire, New Jersey, New York, Washington, and Wisconsin [46]. These states use a combination of encoded first name, last name, middle initial, and date of birth [46]. Maryland uses driver's license numbers as part of its verification check before delivering absentee ballots online to voters, but it is also a state with predictable driver's license based on a voter's last name, first name, middle initial, and birth month and date [19], [20], [47].

4. Access voter registration data

   All states provide copies of voter rolls to political campaigns and others. These data often include the name, address, date of birth, party designation and voting history for each voter. These data do not usually include the SSN or driver's license number of voters, however. Some states provide the data for a fee, while others provide it freely online. For example, anyone can download a copy of the voter registration data for Ohio [48]. Data brokers that specialize in voter data, such as Aristotle [49] and L2 [50], also make voter registries available.

   Once released, a copy of a voter list may be shared widely. For example, a Republican contractor's database that contained nearly every voter in the United States was left exposed online for anyone to copy or access [51].

Audit logs

Webserver logs can record the date, time, and Internet addresses of mobile devices and computers connecting to a website. Database logs can record all changes made to the voter database over time. Maintaining these kinds of audit logs will not prevent a change of address attack or an absentee ballot attack. However, audit logs can help determine, after the fact, whether an attack actually occurred and could possibly provide information about the attacker and the extent and nature of the attack.

The 2014 Pew survey found only 11 out of the 13 states used audit logs for their online voter registration system, with California and Arizona not providing a response to the question [6]. The registrar for the voter data in Riverside, California reportedly maintained no logs [1], [2].

As mentioned earlier, the IRS faces similar cyber-security and identity theft challenges. In response to its challenges, IRS mandates that certain security requirements be met by Authorized IRS e-File Providers [52], such as a requirement to maintain audit logs [52].

CAPTCHAs

An impostor can manually submit address changes on a voter website, one voter at a time. To increase the number of voter record changes per hour, he could employ more people to do the same or automate the process by having a computer program submit the kind of changes at the website that he would submit by hand. Once configured, a computer program could conduct a change of address attack or an absentee ballot attack without human intervention, iteratively impersonating each targeted voter. In addition, a computer program can operate on multiple machines simultaneously. Automation can dramatically increase the number of voter record changes submitted per hour.

A CAPTCHA (Completely Automated Procedures for Telling Computers and Humans Apart) can slow automated attacks. CAPTCHAs are a security scheme first proposed in 2000 and now

widely used online to determine whether a user is more likely to be a human than a computer program [53], [54]. A CAPTCHA displays an image or group of images on the web page and asks the viewer to enter the text displayed in the image or to answer a question about the displayed image(s). Humans can usually respond easily to a CAPTCHA, but computer programs tend to have a difficult time interpreting images, making the proper response to a CAPTCHA more difficult for a program to achieve. Therefore, if a voter website has a CAPTCHA, the attacker has to find a way to automate or semi-automate responses to the CAPTCHA in order for his computer program to iterate its execution over multiple voter records [54], [55].

Figure 1 shows the voter registration pages used by Delaware registered voters to submit address changes. Voters have one of two choices, to either enter name, date of birth, and ZIP (Figure 1a) or enter driver's license number and date of birth (Figure 1b). Both options display a CAPTCHA at the bottom of the page. The voter in the example shown has to enter "LANARK" in the CAPTCHA field to proceed. Visiting the web page again requires the voter to enter different text in the CAPTCHA field.

The IRS in its security requirements for Authorized IRS e-File Providers [52], the NCSL in its advice to states constructing online voter registration systems [4], and the National Institute of Standards and Technology in its Guidelines on Securing Web Servers all recommend that websites that accept information from visitors to the website use CAPTCHAs [56].

As we just described, CAPTCHAs may increase the effort required to automate an attack, but a CAPTCHA does not necessarily slow or prevent automated attacks. As an example, consider a 2010 criminal case in which a hacker reportedly wrote a computer program to impersonate thousands of individual ticket buyers on websites of online ticket vendors such as Ticketmaster, Musictoday and Tickets.com in order to automatically purchase premium event tickets and then resell the tickets later at higher prices [57]. The websites asked for personal financial information for the purchase and used CAPTCHAs to help thwart automated ticket buying [57]. The program defeated the CAPTCHAs presented and ran on a network of computers simultaneously to scale the attack. According to prosecutors, the program grabbed more than 1 million tickets for concerts and sporting events, and the resale of those tickets between 2002 and 2009 yielded more than $25 million in profit [57].

A semi-automated way an attacker's program could defeat CAPTCHAs is to use cheap human labor, such as Amazon Mechanical Turk's Human Intelligence Tasks [58]. The attacker's program electronically routes a copy of the CAPTCHA to a human who views the CAPTCHA and electronically sends the response back to the program, which in turn submits the correct response to the website.

An automated way an attacker's computer program could defeat a CAPTCHA is to use a dictionary of all the CAPTCHA images that could possibly appear on the web page. This approach assumes that the number of CAPTCHA images available at the website is small. The

attacker records each image along with the proper response for each image in a dictionary. Later, when the computer program encounters the CAPTCHA, it looks up the image in its dictionary and responds with the pre-stored answer to respond to CAPTCHAs automatically. The ticket-scalping scheme described above created a dictionary of about a thousand CAPTCHAs found at popular ticket selling websites. Their program then used the dictionary to bypass the CAPTCHAs.

Advanced programming can also defeat some CAPTCHAs straightforwardly by responding to the CAPTCHA as a human would [59], [60], [61]. For example, Google engineers defeated street image CAPTCHAs using a Street View algorithm designed to decipher blurry street addresses [61]. They achieved 99.8 percent accuracy. A company named Vicarious developed software it claims can crack up to 90 percent of CAPTCHAs – including reCAPTCHA – offered by Google, Yahoo and PayPal [60]. Published academic papers use machine learning algorithms to defeat rendered-text CAPTCHAs [59].

Methods

For convenience, we focused on the change of address attack, with the understanding that similar methods apply to the absentee ballot attack and, in some states, to the bulk deletion of voter registrations.

Could attackers have launched a change of address attack on state websites for the 2016 U.S. presidential election? If so, which state websites were likely targets? What personal information did an attacker need? How much would an attack cost?

Here are the six investigative steps we took to answer these questions.

1. Locate state websites.

2. Record what personal information state websites required to identify a voter.

3. Survey sources of personally identifying information.

4. Record and test CAPTCHA defenses of websites.

5. Test automation.

6. Compare total attack costs.

Below we describe these six investigative steps in detail. In the next section, titled Results, we report on our findings for each step and we end this paper with a Discussion section that describes some of the kinds of back office processes and election practices that states may have to detect or limit attacks.

Step 1. Locate state websites.

We surveyed the Internet for state websites that allowed voters to submit changes to their voter registration information online. Websites sought may or may not be the same websites as those that allow members of the public to register to vote, to look up voter information, to review ballot information, or to locate polling places. We searched for a website for each of the 50 states and the District of Columbia. Search queries began with the name of the state followed by "voter registration address change." We reviewed the top 6 results from each search to see whether a candidate website was found.

Step 2. Record what personal information is required at state websites.

For each website found in Step 1, we recorded what personal information the website required from a visitor in order for the visitor to identify himself as a voter and submit changes to the voter's home address.

We advanced through a sequence of one or more web pages the website required, but we stopped before actually changing voter rolls. States expect the general public to use these websites, so the websites tend to provide clear notice at which point a subsequent click will submit a change to voter records. We only proceeded in the sequence of web pages as we were confident that doing so does not submit any changes to actual voter rolls. As we proceeded through the pages, we recorded the fields of personal information that the website required.

Because we did not actually change any records, we may have underreported the required fields. For example, a website could request additional fields of personal information on subsequent pages viewed only if the visitor proceeds beyond where we stopped. Therefore, our approach provides the minimum personally identifying information required by each website.

Our recorded fields may not necessarily be all the fields requested on the page. We consider the required fields to be those fields necessary to proceed; these are usually highlighted, stated as such on the web page, or become evident if we attempt to submit a blank page. For example, a web page may request Social Security number, but if the value was not required to proceed, we did not record it as being a required field.

A field of personal information may be required but not actually checked against any information the state holds. For example, a website may require a Social Security number be given, but behind the scenes, the value given is not actually checked for validity in order to submit the change; in this case, any value having 9 digits would suffice without necessarily being the SSN of the voter. Because we were not actually changing any records, we had no way of knowing which values the state actually verified. Therefore, our results may report a greater burden for the impostor than is actually necessary to perpetrate the attack.

Lastly, the website may not require the same fields that a Frequently Asked Questions (FAQ) page for the website describes. In this case, we deferred to the version available on the

website itself unless it had a multi-page process, in which case, we elected to report what the FAQ described in the belief that there may be other information requested on subsequent pages we could not view without risk of changing a voter record.

We made a screen shot of each page encountered for each website. If there were options that could lead to more than one template of fields, we recorded all accessible paths.

Step 3. Survey sources of personally identifying information.

We surveyed publicly available sources of voter records, Social Security numbers, and driver's license numbers that the attacker would need to impersonate voters on the state websites. We identified data sources, assessed data quality and access issues, and recorded the cost of acquisition.

An attacker looking to impact a local election only needs data specific to the targeted geography. An attacker looking to cause widespread havoc on voter rolls might garner as much data as practical to impact as many state websites as possible. Foreign governments or state actors may simply steal the data needed. The specific data and amount of data needed by an attacker depends on the voter websites targeted. An attacker could pursue several different means of acquiring the personal voter data needed.

Sources of personal information about American voters are the government, data brokers, and online outlets from anonymous sources. Figure 2 summarizes ways an attacker could acquire needed personal information. The leftmost path describes an overall process of acquiring voter data, adding date of birth information and gender to the voter data as needed, and then using the jointly acquired demographic data to acquire associated Social Security numbers and driver's license data.

An alternative appears on the right side of Figure 2, in which the attacker acquires voter data or personal profiles (data containing names, addresses, and demographics), along with accompanying Social Security numbers and driver's license data, through anonymous sources online. Breached data and data of unknown origins are often available for sale anonymously on the deep web and the darknet. The "deep web" is part of the World Wide Web whose contents are not indexed by standard search engines. Web searches tend to exclude content from the deep web, making it harder for someone to just stumble onto the content. A user can access content freely on the deep web if she knows (or learns) the URL. The "darknet" is a specialized encrypted network of anonymity-providing web servers on the deep web that primarily operate as a haven for drug markets, pedophiles, sex traffickers, and personal data [62]. An attacker could also mix named and anonymous sources.

We surveyed the availability of the personal data needed for each of the 36 websites, as available from these different kinds of named and anonymous sources and explored ways to combine them.

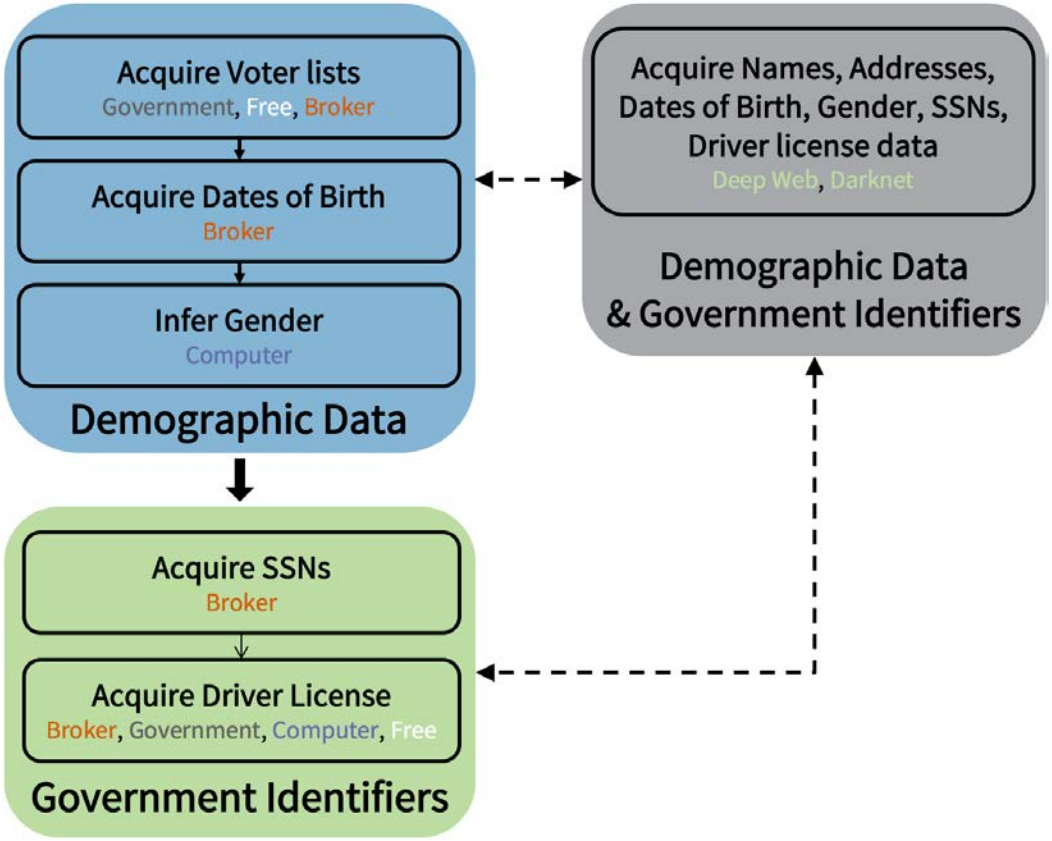


Figure 2. Ways an attacker can acquire the personal data needed to impersonate voters. On the left, top to bottom, is a sequential build-up of demographic information (including date of birth, or DOB) and government-issued identifiers (namely, Social Security numbers, or SSNs, and driver's license) from named sources. Different named sources include government offices, free downloads, data brokers, and computer programs. On the right is an alternative that acquires the data online through unknown sources on the deep web or darknet. Additional options are combinations of named and unnamed sources between the left and right sides (connected by dashed lines).

Step 4. Record and test CAPTCHA defenses of websites.

For each website found in (1), we recorded whether the sequence of web pages we visited contained a CAPTCHA, and if so, what kind of CAPTCHA appeared: text image, street image, or collection of images. Figures 1 and 3b show examples of street image CAPTCHAs. Figure 3a shows a text image CAPTCHA, and Figures 3c and 3d show CAPTCHAs that ask a question based on a collection of images.

If a CAPTCHA was present on the sequence of web pages we visit, we performed the following tests to help determine whether automated or semi-automated responses to the CAPTCHA were feasible.

We captured 100 instances of the CAPTCHA from each of the websites to understand the nature of the images used. We then surveyed to see whether computer programs existed that would provide proper responses to the kind of CAPTCHA displayed or whether simple computer programs could have been written to defeat the CAPTCHA.

Our test for semi-automation assumes that the CAPTCHA can be re-sent by a computer program to an online network of human helpers, who view the CAPTCHA and then send the response back to the program to submit.

For each of the websites having a CAPTCHA, we entered data in all required fields except the CAPTCHA field, and then left the web page idle for 2 and again for 5 minutes. If the web page just waited for a response and then accepted a response to the CAPTCHA after the 2 or 5 minutes expire, then we considered the website ripe for semi-automation, because 5 minutes would be more than enough time to get an answer to the CAPTCHA from a network of human helpers. On the other hand, if during the 2 minutes the web page refreshed the CAPTCHA to displayed another image, or terminated the session altogether, we considered the website resilient to semi-automated responses to the CAPTCHA.

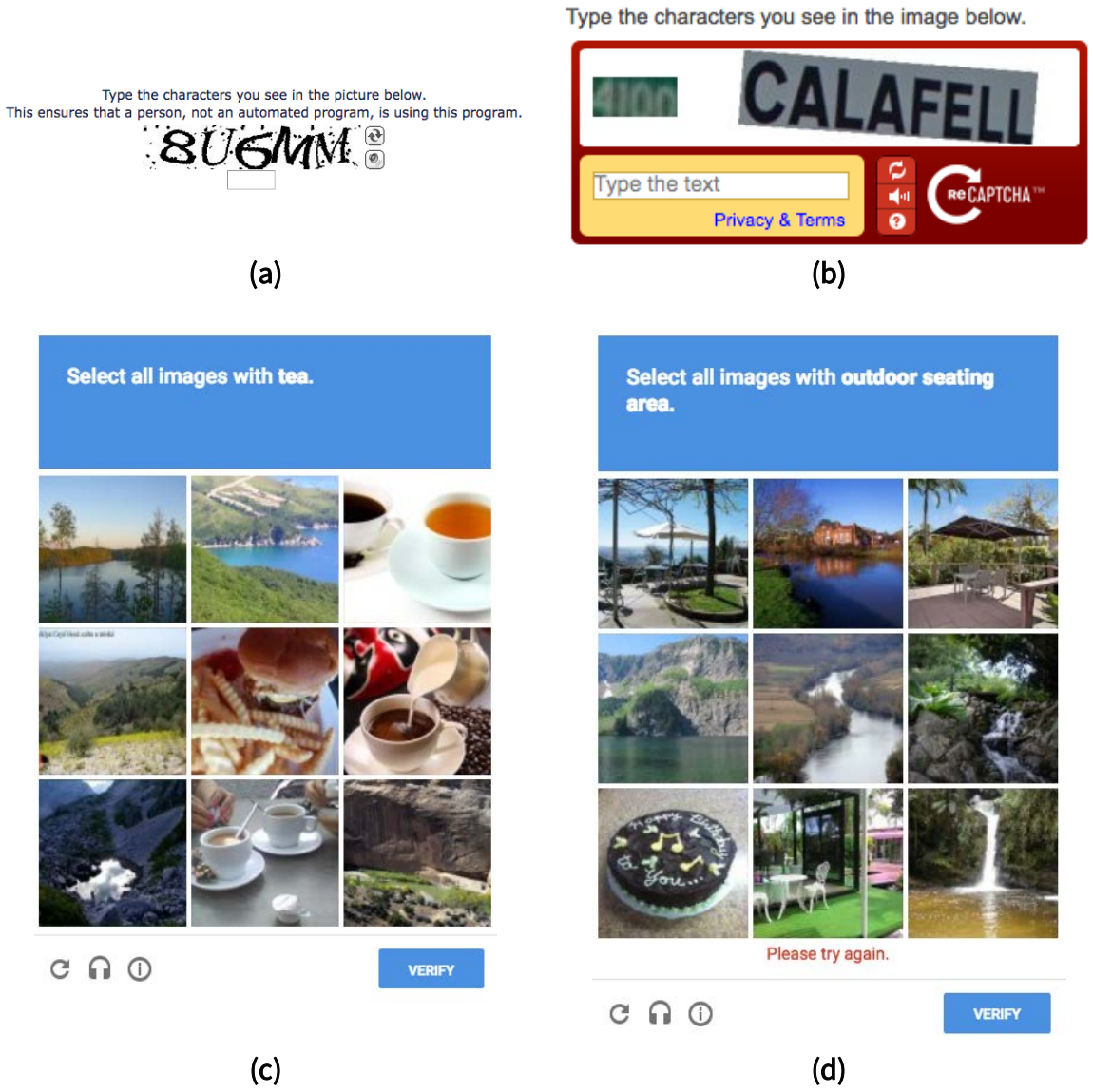


Figure 3. Different kinds of CAPTCHAs used on state voter websites: (a) text CAPTCHA from Ohio's website, (b) street image CAPTCHA from Indiana's website, (c) and (d) collection of image CAPTCHAs from Nebraska's website. See Figure 4 for the URLs for each state.

Step 5. Test automation.

We constructed programs capable of automatically entering personal voter information into state websites to locate polling places. This allowed us to demonstrate the effort involved in coding a program that uses personally identifying information on voters and that continued across multiple web pages. We used polling place websites to avoid making any changes to the content of actual voter records.

Step 6. Compare total attack costs.

In this step, we compared the relative costs of implementing the change of address attack on state voter websites based on information and facts acquired in the prior steps. Personally-identifying information has acquisition costs. Techniques to defeat CAPTCHAs have varying programming costs and may also have human costs if semi-automated. We itemized these and other costs per state to compute the "total attack cost" for the change of address attack. Here are the methods used:

Step 6a. Total Attack Costs Using Named Sources. Compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using personal data from named sources with automated means to defeat CAPTCHAs.

Step 6b. Total Attack Costs Using Darknet Sources. Compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using personal data primarily from darknet sources with automated means to defeat CAPTCHAs.

Step 6c. Total Attack Costs Using Deep Web Sources. Compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using personal data from deep web sources with automated means to defeat CAPTCHAs.

Step 6d. Total Attack Costs Using Semi-Automated Means to Defeat CAPTCHAs. Compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using a network of human helpers to defeat CAPTCHAs.

Step 6e. Comparison of Total Attack Costs by State. Compute total attack costs by state as a means of assessing relative risk.

Results

Below is a summary of our findings. Supporting details appear in the subsections that follow this summary and are organized following the 6-step approach described above.

We found online voter websites for 35 states and D.C. that allowed voters in those states to submit changes to their addresses. The kinds of personal information an attacker would have provided were the voter's name in 30 (83 percent) of 36 websites, date of birth (35 or 97 percent), driver's license number, or State ID if the voter did not drive (33 or 92 percent), and part or all of the voter's SSN (22 or 61 percent). There were 43 possible combinations of fields of personal voter information required at the 36 websites. In combination, {name, date of birth, driver's license} was needed in 27 (63 percent) of the 43 and {name, date of birth, address} was necessary in 15 (35 percent).

Our surveys found that an attacker could acquire voter names and demographics from government offices, data brokers, the deep web, or darknet markets, and could acquire

government-issued identification numbers from data brokers, government sources, websites, and the deep web or darknet markets.

For voter lists, we calculated that an attacker could have spent $219 to acquire voter lists for 18 (50 percent) states, $4,407 for 29 (81 percent) states, or $17,679 for all 35 states and D.C. from authoritative government sources, publicized websites, or data brokers. Sources of all cost information and the assumptions we used to calculate attack costs are laid out further below in the sub-sections.

Alternatively, an attacker could have spent $1,002 from darknet sources to acquire 2 datasets that jointly contained the names, addresses, dates of birth, genders, and SSNs of most adult Americans.

While prices varied, an attacker could spend as little as $40 per month for an unlimited number of searches at a data broker website to acquire the SSN and driver's license numbers of voters (and an additional $0.01 per record for more details such as prior addresses and driver's license issue date). Some data brokers charged $1 per name compared to $0.41 from clandestine sources on darknet markets. Compilations from swipes of the magnetic strip of driver's licenses and photo images of driver's licenses were available on darknet markets for prices ranging from $0.01 to $15 each.

Government election offices, political campaigns, and data analytic companies working on elections have experienced breaches containing most or all the personal information an attacker would have needed to impersonate voters at the 36 websites. Copies of some of these datasets have been found publicly available on the deep web and could have been available for an attacker to use if the attacker made or became aware of the URLs. In these cases, the data would have been available at no additional cost.

Eleven (31 percent) of the 36 websites had a CAPTCHA service that attempted to limit the speed with which an attacker could submit address changes on the website. In 2016, however, automated programs could respond to the kinds of CAPTCHAs found on all the state websites that had CAPTCHAs, thereby rendering them a nominal deterrent.

In summary, the cost of an attack may vary based on actual data sources and whether the website used CAPTCHAs. An attacker that primarily used two datasets offered on darknet markets, the Texas voter list, and a data broker as sources could change 1 percent of the voter records on all 36 websites for a total attack cost of $10,081. The minimum state cost is $1 (Alaska), the maximum $3,059 (Texas), and the median $41 per state. Alternatively, an attacker could primarily purchase data from government offices, data brokers and websites archived on search engines to change 1 percent of the voter records on all 36 websites for $24,926. The minimum state cost was $5 (Delaware), the maximum $3,059 (Texas), and the median $417 per state. An attacker that found relevant information on the deep web, or had a confederate that placed it there, would have no data costs, dramatically reducing the attack cost for machine use to $748 for all 36 state websites.

Results for Step 1: Number of Websites

Thirty-five states and D.C. offer online websites for submitting changes to voter addresses. Specifics appear below.

From June 29 to July 6, 2016, and again in November 2016, we searched google.com for official state election websites. Our search queries were of the form "<state> voter registration address change," where <state> is the name of a state. An example is "Alabama voter registration address change" for the state of Alabama. We reviewed the top 6 results from each search. Using this approach, we located 35 websites for states and one website for the District of Columbia (D.C.). All 36 websites allowed a voter to submit online changes to a residential address in order to subsequently change the address in the voter rolls. The 35 states were: Alabama, Alaska, Arizona, California, Colorado, Connecticut, Delaware, Georgia, Hawaii, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, Nevada, New Jersey, New Mexico, New York, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Texas, Utah, Vermont, Virginia, Washington, and West Virginia. Figure 4 lists these states and the URLs of their websites. Figure 5 shows them geographically.

Thirty (30 or 83 percent) of the 36 websites allowed a registered voter to submit a change of address at the same website he would register to vote. Some of these may require the change of address to be done by re-registering. The other six states—Colorado, Michigan, New Jersey, Ohio, Texas, and the District of Columbia—provided different websites for a registered voter wanting to change an address online.

South Carolina required a voter to make changes on two websites in order to submit a residential address change online. A voter would first submit an address change to the voter's driver's license information at a motor vehicle website, and then submit a change of the address to the voter registration record at a voter registration website.

We did not find an online change of address option for 15 (or 29 percent) of the 51 possibilities (50 states and D.C.). Specifically, Arkansas, Florida, Idaho, Iowa, Maine, Mississippi, Montana, New Hampshire, North Carolina, North Dakota, Oklahoma, South Dakota, Tennessee, Wisconsin, and Wyoming did not have online options for submitting voter address changes. Most of these states had an online web page that specifically described offering mail-in or in-person options only.

State	State Website for Changing Voter Information Online	Same as Voter Registration website?
Alabama	https://www.alabamavotes.gov/olvr/default.aspx	Y
Alaska	https://voterregistration.alaska.gov/	Y
Arizona	http://servicearizona.com/voterRegistration?popularclick	Y

State	State Website for Changing Voter Information Online	Same as Voter Registration website?
Arkansas		
California	http://registertovote.ca.gov/	Y
Colorado	https://www.sos.state.co.us/voter-classic/pages/pub/olvr/findVoterReg.xhtml	N
Connecticut	https://voterregistration.ct.gov/OLVR/registerDetails.do	Y
Delaware	https://ivote.de.gov/voterlogin.aspx	Y
Florida		
Georgia	https://registertovote.sos.ga.gov/GAOLVR/welcome.do	Y
Hawaii	https://olvr.hawaii.gov/	Y
Idaho		
Illinois	https://ova.elections.il.gov/	Y
Indiana	https://indianavoters.in.gov/publicsite/ovr/introduction.aspx	Y
Iowa		
Kansas	https://www.kdor.ks.gov/Apps/VoterReg/Default.aspx	Y
Kentucky	https://vrsws.sos.ky.gov/ovrweb/	Y
Louisiana	http://www.sos.la.gov/electionsandvoting/pages/onlinevoterregistration.aspx	Y
Maine		
Maryland	https://voterservices.elections.state.md.us/OnlineVoterRegistration/InstructionsStep1	Y
Massachusetts	https://www.sec.state.ma.us/ovr/	Y
Michigan	https://onlineservices.michigan.gov/ExpressSOS/CustomerManagement/EligibilityRequirements	N
Minnesota	https://mnvotes.sos.state.mn.us/VoterRegistration/VoterRegistrationMain.aspx	Y
Mississippi		
Missouri	https://www.sos.mo.gov/votemissouri/request	Y
Montana		
Nebraska	https://www.nebraska.gov/apps-sos-voter-registration/	Y
Nevada	https://nvsos.gov/sosvoterservices/Registration/step1.aspx	Y
New Hampshire		
New Jersey	https://emvc.state.nj.us/MVC_SAC/AVSStart.jsp	N
New Mexico	https://portal.sos.state.nm.us/OVR/WebPages/InstructionsStep1.aspx	Y
New York	https://dmv.ny.gov/more-info/electronic-voter-registration-application	Y
North Carolina		

State	State Website for Changing Voter Information Online	Same as Voter Registration website?
North Dakota		
Ohio	https://olvr.sos.state.oh.us/ovru/Modify.aspx	N
Oklahoma		
Oregon	https://secure.sos.state.or.us/orestar/vr/register.do	Y
Pennsylvania	https://www.pavoterservices.state.pa.us/Pages/VoterRegistrationApplication.aspx	Y
Rhode Island	https://vote.sos.ri.gov/	Y
South Carolina	https://www.scdmvonline.com/dmvpublic/trans/addrchange.aspx then https://info.scvotes.sc.gov/eng/ovr/start.aspx	Y*
South Dakota		
Tennessee		
Texas	https://txapps.texas.gov/tolapp/sos/SOSACManager	N
Utah	https://secure.utah.gov/voterreg/index.html	Y
Vermont	https://olvr.sec.state.vt.us/	Y
Virginia	https://vote.elections.virginia.gov/VoterInformation	Y
Washington	https://weiapplets.sos.wa.gov/myvoteOLVR/myvote/UpdateAddressWelcome?voterid=WA002233217&language=EN&org=	Y
Washington, DC	https://www.vote4dc.com/UpdateApply/SearchOldVoter	N
West Virginia	https://ovr.sos.wv.gov/Register	Y
Wisconsin		
Wyoming		

Figure 4. The website URLs for 35 states and D.C. (jointly termed "36 states") that allowed visitors to update voter addresses online in 2016. References in this writing to 36 state websites include Washington, D.C., which is not a state but is also not part of any state. All but 6 of the 36 websites used the same website for both voter address changes and voter registration. *South Carolina required changing a voter address at two websites, one a voter registration website. We count South Carolina once in references to the total number of states or state websites.

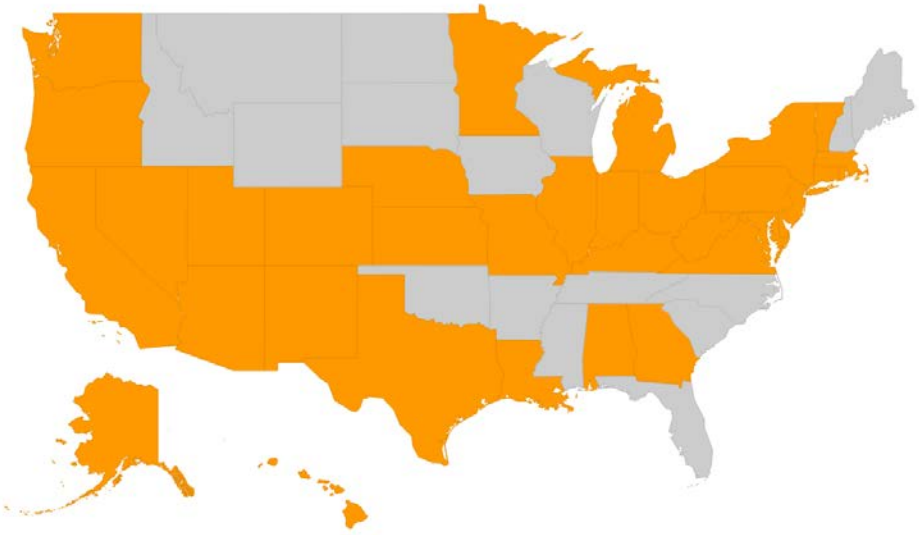


Figure 5. Map of the United States, color-coded to distinguish the 35 states and the District of Columbia (jointly termed "36 states") that provided websites to change voter addresses online (orange) from the states that provided no online means for changing voter information (gray).

Results for Step 2: Personal Information Required

Submitting voter address changes online required providing date of birth at 35 (97 percent) of 36 websites, driver's license number, or state ID number if the voter did not drive, at 33 (92 percent) of 36 websites, name in 30 (83 percent), part or all of the voter's SSN at 22 (61 percent), address information at 15 (42 percent), and gender at 7 (19 percent). There were 43 possible ways personal voter information could be entered into the 36 websites. The voter's name, date of birth, and driver's license was needed in 27 (63 percent) of the 43 possible ways. The voter's name, date of birth and address was necessary in 15 (35 percent) of the 43. Nineteen (44 percent) of the 43 possible ways included the voter's name, date of birth and the last four digits or all of the voter's SSN. Details appear below.

A visitor to any of the websites had to provide some combination of personal information about the voter before submitting a change of the voter's address. For example, on Delaware's website, a visitor could choose to provide either: (1) the name, date of birth, and ZIP (postal) code of the voter (Figure 1a); or (2) date of birth and driver's license number of the voter (Figure 1b). On Alabama's website, a visitor must provide the voter's name, date of birth and driver's license number or state ID if the person does not drive (Figure 6). South Carolina's website additionally required SSN, gender and race of the voter; see Figure 7.

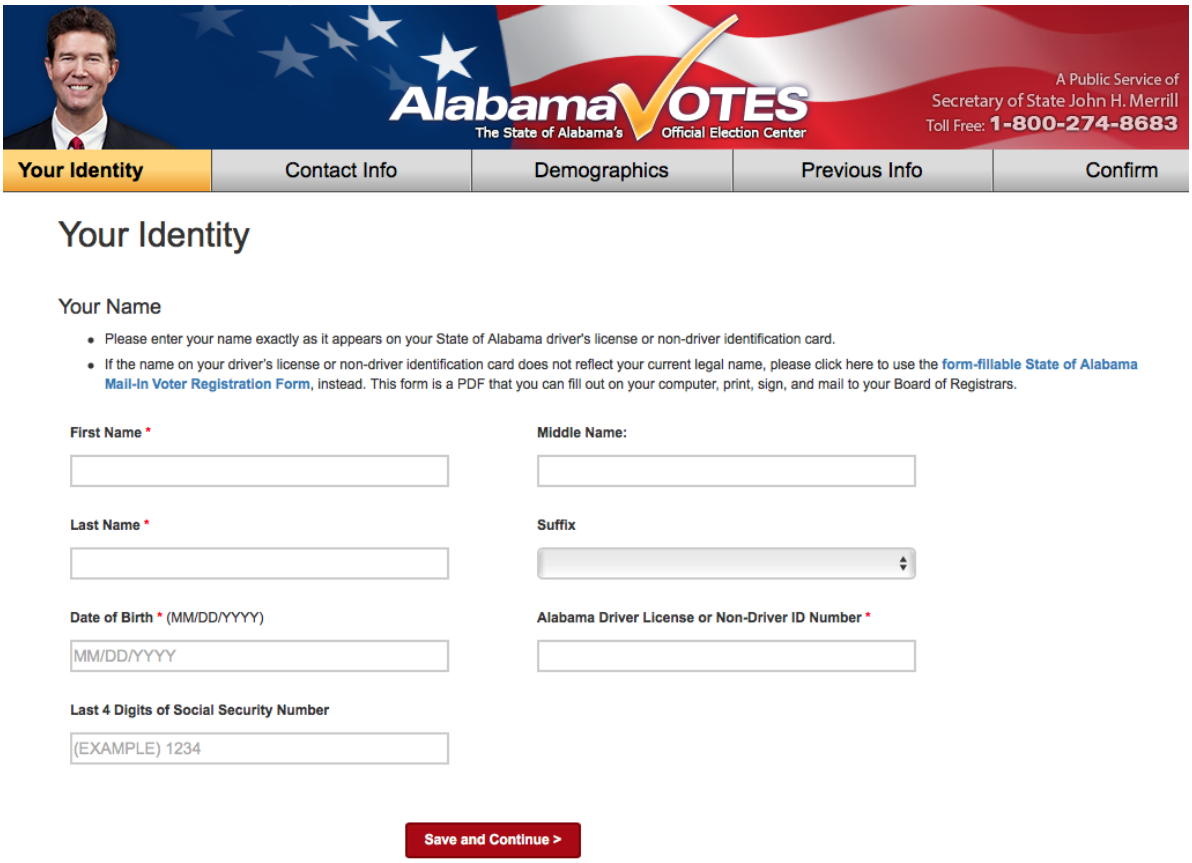


Figure 6. Information required to authenticate a voter at the Alabama website was: name, date of birth, and driver's license number (or state ID for a non-driver). Required fields have red asterisk (*), so a visitor may optionally enter the last 4 digits of his SSN. See Figure 4 for the source URL.

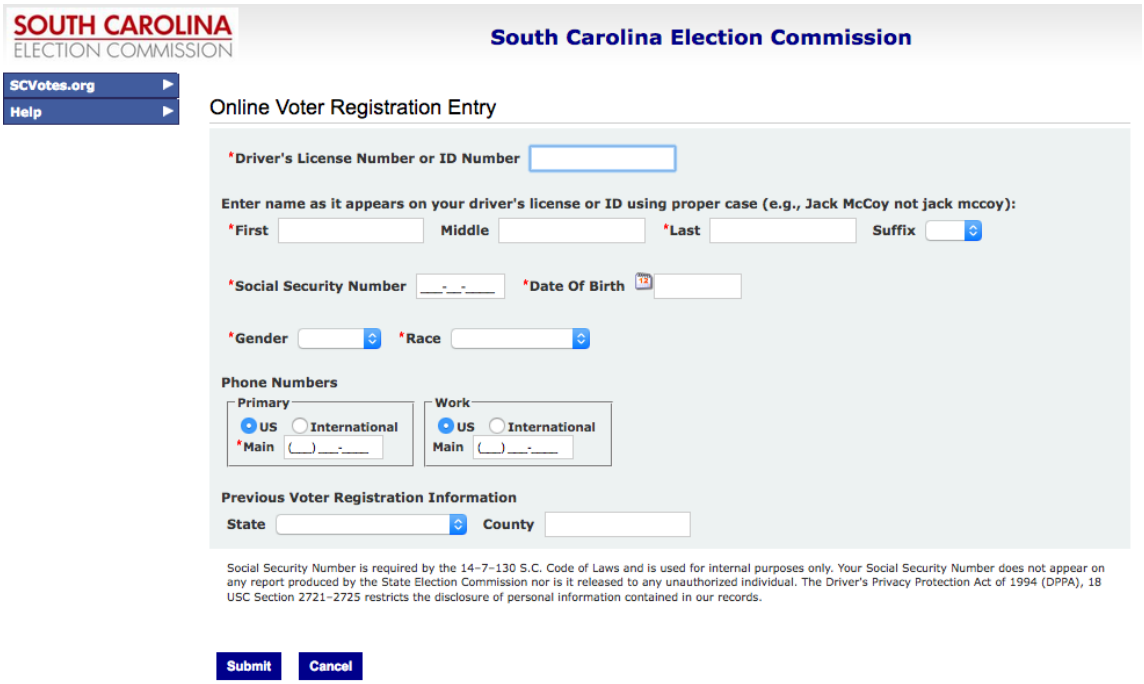


Figure 7. Information required of a voter to identify himself at the South Carolina website was: driver's license number (or state ID if he does not drive), name, SSN, date of birth, gender, race and a telephone number. See Figure 4 for the source URL.

Table 2 lists the minimal personal information demanded by each of the 36 websites, and Table 3 enumerates the information by field. We report the minimal information because, as we described our methodology earlier, we stop surveying information at the point we believe any further advance on the website may alter a voter's data. An attacker may reasonably conclude these are the required fields because of website design, notwithstanding discrepancies between these and state FAQs, as described in our Methods section earlier.

Name and Demographic Information. The 36 state websites relied heavily on name and demographic fields. The voter's first and last name was required in 30 (83 percent) of the 36 websites, date of birth in 35 (97 percent) of the websites, and some part of the voter's address in 15 (42 percent) of the websites. New Jersey had the only website that did not require the voter's date of birth. The South Carolina website was the only website that required race.

The Delaware website requested a visitor to enter the voter's name, date of birth and 5-digit ZIP code; see Figure 1. The Connecticut website required a visitor to enter the voter's name, date of birth, and town if the voter did not have a driver's license.

Websites for Minnesota, New York, and South Carolina requested a telephone number or an email address. Does the voting office have a reliable stored value of a telephone number or an email address for comparison? If not, values entered for phone and email are not verified against any previously stored values, but instead may be used for future verification of the

voter or as a means for subsequent contact if needed. Therefore, values entered on the website for these fields would not have to be those actually held by the voter. A made-up value or a non-traceable phone number, such as one purchased on Skype [63], or a free email account, such as would be available on Gmail [64], could suffice. See Figure 7 for an example.

Missouri's website had asterisks on the required fields {name, date of birth, address, and last 4 digits of SSN}. Driver's license number and gender were among the fields listed, but not marked as required. When we attempted to submit a blank page, a message appeared that listed the fields required to complete, which were the same as those marked with asterisks but also included gender. Driver's license number, a field like gender that did not have an asterisk, was not listed as being required. For this reason, we include gender as a required field for Missouri. We did not submit any actual changes, so we do not know whether Missouri's website actually required gender. However, it seems plausible, that under these circumstances, an attacker might believe gender was required. If gender was not required, our assumption unduly increased the burden on the attacker.

Government-Issued Numbers. The 36 state websites also relied heavily on government-issued number—namely, SSNs and driver's license numbers. Thirty-three (92 percent) of 36 websites requested a driver's license number. If a voter did not have a driver's license, a voter's state ID issued by the same state department that regulates motor vehicle driver's licenses could be entered; so, for convenience in this writing, all subsequent references to "driver's license" include non-driving state IDs issued by the same state department that issues driver's licenses in the state unless stated otherwise or obvious from context.

Of the 33 websites that requested a driver's license number, only the websites for Colorado and Connecticut gave visitors the option of alternatively providing the last 4 digits of the voter's SSN or using the voter's name, date of birth and ZIP code, respectively. These are denoted by "OR" on Table 2. If a voter did not have a driver's license number (or state ID number), a visitor could enter the last 4 digits of the voter's SSN in Connecticut, Minnesota, Pennsylvania, Vermont, and District of Columbia websites. These options are listed as "ELSE" on Table 2. In all other cases, if a voter could not provide a requested driver's license, a mail-in or in-person option was necessary.

Twenty-two of the 36 (61 percent) websites requested all or part of the voter's SSN. Five (14 percent) websites required the full 9-digit SSN, while 17 (47 percent) required only the last 4 digits. Only the voter's date of birth and SSN was necessary for Kentucky's website.

Connecticut and Delaware websites (see Figure 1) allowed a visitor to enter combinations of fields of personal information that did not include any government-issued ID number. All the other states and D.C. required a government-issued ID number, though none of the websites relied solely on a government-issued ID number. Fourteen (39 percent) websites accepted combinations that included both a driver's license number and some form of SSN, 19 (53

percent) accepted combinations with driver's license and no SSN, and 8 states (22 percent) accepted combinations with some form of SSN and no driver's license number.

Other Kinds of Information. California and Pennsylvania websites required party affiliation. Illinois, South Carolina, Washington required the issuance date of the driver's license number. New York and Texas required document numbers that appear on the driver's license. Michigan required eye color, which also appears on the driver's license. Texas required a voter identification number. And Kansas required copies of citizenship documents on initial voter registrations but not on changes to existing registrations.

Combinations of Fields of Personal Information. As mentioned earlier, the websites for Colorado, Connecticut, Delaware, Minnesota, Pennsylvania, Vermont, and Washington, D.C. allowed a visitor to enter one of two different combinations of fields of personal voter information. Therefore, there were 43 possible combinations of personal voter information to enter for the 36 websites.

Table 2 identifies the 43 combinations of fields of personal voter information necessary at the 36 state websites. For those websites that offered a choice, Table 2 describes whether the choice was an OR, in which case a visitor could select either combination of fields to enter, or was an ELSE, in which case a visitor could only use the second choice if the voter did not have a driver's license number (or non-driving state ID number). If a state website required information such as party affiliation or voter identification number, Table 2 identifies that "Other" information was required. If the website required any information that could be found on the driver's license, such as issuance date, audit number, or eye color, Table 2 identifies that "Driver's License Other Info" was required.

Table 4 lists which fields of voter information were required at the same state websites. Most websites (16) required a unique combination of fields of personal information. The voter's name, date of birth, and driver's license number were required at 8 of the websites. Five websites additionally required the voter's address, and, alternatively, five websites additionally required the last four digits of the voter's SSN. See Table 4 for a complete list.

Overall, the voter's name, date of birth, and driver's license were needed in 27 (or 63 percent) of the 43 possible ways personal voter information could be entered into the 36 websites. The voter's name, date of birth, and address were necessary in 15 (35 percent) of the 43 possible combinations. Nineteen (44 percent) of the 43 possible ways included the voter's name, date of birth and the last four digits or all of the voter's SSN.

State	ID requirements
Alabama	Name, DOB & DL number
Alaska	Name, DOB, DL number, & Last 4 SSN
Arizona	Name, DOB, DL number & Last 4 SSN
California	Name, DOB, Street, ZIP, County, Party, Last4 SSN & DL number

State	ID requirements
Colorado	Name, DOB, ZIP & DL number OR Name, DOB, ZIP & Last4 SSN
Connecticut	Name, DOB, Town & DL number ELSE Name, DOB & Town
Delaware	Name, DOB & Zip code OR DOB & DL number
Georgia	Name, DOB, County & DL number
Hawaii	Name, DOB, Gender, SSN & DL number
Illinois	DOB, DL number & DL issue date
Indiana	Name, DOB, DL number
Kansas	Name, DOB, DL number & [Citizenship documents]
Kentucky	DOB & SSN
Louisiana	Name, DOB, Sex & DL number
Maryland	Name, DOB, Last 4 SSN & DL number
Massachusetts	Name, DOB & DL number
Michigan	Name, DOB, DL number, Eye color & Last4 SSN
Minnesota	Name, DOB, DL number & Email ELSE Name, DOB, Last4 SSN & Email
Missouri	Name, DOB, *Sex, Address & Last 4 SSN
Nebraska	Name, DOB & DL number
Nevada	Name, DOB, ZIP, DL number & Last4 SSN
New Jersey	ZIP, DL number & SSN
New Mexico	Name, DOB, Sex, DL number & SSN
New York	DOB, ZIP, DL number, DL doc number, last 4 SSN & Email
Ohio	Name, DOB, DL number & last 4 SSN
Oregon	Name, DOB & DL number
Pennsylvania	Name, DOB, Address, Party & DL number ELSE Name, DOB, Address, Party & Last4 SSN
Rhode Island	Name, DOB, ZIP & DL number
South Carolina	DL number, DL issue date & SSN THEN Name, DOB, Sex, DL number, SSN, Race & Phone
Texas	DOB, County, DL number, DL audit number, Last 4 SSN & Voter ID number
Utah	DOB, Street number & DL number
Vermont	Name, DOB, Town, DL number ELSE Name, DOB, Town & Last 4 SSN
Virginia	Name, DOB, County & Last 4 SSN
Washington	Name, DOB, Sex, DL number & DL issue date
Washington, DC	Name, DOB, DL number ELSE Name, DOB & Last4 SSN
West Virginia	Name, DOB, DL number & Last 4 SSN

Table 2. Information required from visitor at a state website to be accepted as a voter allowed to change the voter's address, based on an attacker's inspection of the website. Fields included date of birth (DOB), driver's license or state identification ID numbers (DL

number), and other information combined with the DL number, gender (Sex), and Social Security numbers (SSN), including just the last 4 digits (Last 4 SSN). CT and other states offered a choice between two groups of fields to provide ("OR"). SC required visiting two websites ("THEN"). CT, DC, MN, PA, and VT allowed the second choice of fields only if the visitor did not have a DL number ("ELSE"). Figure 4 lists the source URLS for each state website. Kansas requires a copy of citizenship documents on an initial registration but not on address changes. *We consider gender as required for Missouri because upon a blank submission, gender was among the list of required fields even though it was not marked as required on the form.

	Name and Demographics				Government Issued Numbers				Other
	Name	Date of Birth	Gender	Address (all or part)	Driver's License / State ID Number	Driver's License Other Info	Full SSN	Last 4 SSN	
Alabama	■	■			■				
Alaska	■	■			■			■	
Arizona	■	■			■			■	
California	■	■		■	■			■	■
Colorado 1	■	■		■	■				
Colorado 2	■	■		■				■	
Connecticut 1	■	■		■	■				
Connecticut 2	■	■		■					
Delaware 1	■	■		■					
Delaware 2		■			■				
Georgia	■	■		■	■				
Hawaii	■	■	■		■		■		
Illinois		■			■	■			
Indiana	■	■			■				
Kansas	■	■			■				
Kentucky		■					■		
Louisiana	■	■	■		■				

	Name and Demographics				Government Issued Numbers				Other
	Name	Date of Birth	Gender	Address (all or part)	Driver's License / State ID Number	Driver's License Other Info	Full SSN	Last 4 SSN	
Maryland									
Massachusetts									
Michigan									
Minnesota 1									
Minnesota 2									
Missouri									
Nebraska									
Nevada									
New Jersey									
New Mexico									
New York									
Ohio									
Oregon									
Pennsylvania 1									
Pennsylvania 2									
Rhode Island									
South Carolina									
Texas									
Utah									
Vermont 1									
Vermont 2									
Virginia									
Washington									
1									

	Name and Demographics				Government Issued Numbers				Other
	Name	Date of Birth	Gender	Address (all or part)	Driver's License / State ID Number	Driver's License Other Info	Full SSN	Last 4 SSN	
Washington, DC 2									
West Virginia									

Table 3. Kinds of information required in order to submit updates online by state. Orange represents a required field, based on an attacker's inspection of the website. Some state websites provided a choice between the sets of fields required, identified as 1 and 2. Table 2 lists the actual fields minimally required by each state's website and for those state websites providing a choice, the OR or ELSE relationship between choices.

State	Required Fields of Personal Voter Information
Alabama	Name, Date of birth, Driver's license number
Indiana	
Kansas	
Massachusetts	
Minnesota 1	
Nebraska	
Oregon	
Washington, DC 1	
Colorado 1	Name, Date of Birth, Address, Driver's license number
Connecticut 1	
Georgia	
Rhode Island	
Vermont 1	
Alaska	Name, Date of birth, Driver's license number, Last 4 SSN
Arizona	
Maryland	
Ohio	
West Virginia	
Colorado 2	Name, Date of birth, Address, Last 4 SSN
Vermont 2	
Virginia	
Minnesota 2	Name, Date of birth, Last 4 SSN

State	Required Fields of Personal Voter Information
Washington, DC 2	
Hawaii	Name, Date of birth, Gender, Driver's license number, SSN
New Mexico	
Connecticut 2	Name, Date of birth, Address
Delaware 1	
California	Name, Date of birth, Address, Driver's license number, Last 4 SSN, Party
Delaware 2	Date of birth, Driver's license number
Illinois	Date of birth, Driver's license number, Driver's license other
Kentucky	Date of birth, SSN
Louisiana	Name, Date of birth, Gender, Driver's license number
Michigan	Name, Date of birth, Driver's license number, Driver's license other, Last 4 SSN
Missouri	Name, Date of birth, Gender, Address, Last 4 SSN
Nevada	Name, Date of birth, Address, Driver's license number, Last 4 SSN
New Jersey	Address, Driver's license number, SSN
New York	Date of birth, Address, Driver's license number, Driver's license other, Last 4 SSN
Pennsylvania 1	Name, Date of birth, Address, Driver's license number, Party
Pennsylvania 2	Name, Date of birth, Address, Last 4 SSN, Party
South Carolina	Name, Date of birth, Gender, Driver's license number, Driver's license other, SSN, Race
Texas	Date of birth, Address, Driver's license number, Driver's license other, Last 4 SSN, Voter ID
Utah	Date of birth, Address, Driver's license number
Washington	Name, Date of birth, Gender, Driver's license number, Driver's license other

Table 4. Groups of states whose websites have the same minimal requirements for visitors to identify themselves as voters. Table 3 lists these requirements by state.

Results for Step 3: Data Sources of Personal Voter Information

In summary, an attacker could have acquired the personal information needed for any or all of the 36 websites from named sources or from clandestine alternatives (Figure 2). We surveyed these two tracks separately. Below is a summary with details in the sub-sections that follow this summary.

When using named sources, the attacker would have started with voter data because it contains a list of actual voters. Starting with a dataset of the names and demographics of the general adult population is not as efficient. The attacker could have acquired voter data for all 36 states from government offices, voting list brokers, or publicized websites. For the 18 (50 percent) states having the least expensive voter data, the total cost would have been $219; for 29 (81 percent) states, the cost would have been $4,407, or for all 36 states, $17,679.

Voter names and all the demographic information needed to impersonate voters at more than 20 (56 percent) of the 36 websites were contained in voter lists acquired from government sources or from voter list brokers. Date of birth information was missing for 13 states in data acquired from government sources and for 11 states in data from voting list brokers. But for less than $30, an attacker could use public data broker websites to append the missing date of birth information. Gender was missing in a Missouri dataset acquired from a voting list broker, but it could have been inferred from the first names of voters or guessed.

SSNs were required on 17 (47 percent) of the 36 websites. Twenty-two data brokers proffered a service that provides an SSN in response to a search of a voter's name and address or to a search additionally including the voter's birth date. While prices varied, the least expensive offer was for an unlimited number of searches at a monthly fee of $40.

Driver's license numbers and sometimes issuance dates were required by 33 (92 percent) of the 36 websites. Eleven data brokers provided this information in response to a search of a voter's name and address or date of birth. Monthly subscriptions ranged from $39.95 to $59.95 per month, and some websites charged an additional $1, $5 or $10 per lookup or 1 cent more for issuance date. Alternatives were found for some states. A Nebraska government website revealed a voter's driver's license number if given the voter's SSN. Driver's license numbers for Illinois, Maryland, Michigan, New Jersey and Washington could be computed directly from voter demographics. Arizona used SSN as a driver's license number for some of its population.

Offers from clandestine sources on darknet markets included access to voter lists or similar population lists for as little as $1 per million and to datasets having names, address, dates of birth, SSNs, and driver's license numbers with issuance dates for as little as $0.41 per record for Americans in all 50 states and the District of Columbia. One-time flat fee offers included $500 for a dataset that reportedly contained the names, address, dates of birth, and SSNs of all adult Americans and $10 for a URL link to a data broker or website having equivalent information. Negotiable volume discounts beyond these advertised prices appeared widely.

National and statewide voter datasets, as well as the SSNs and driver's license data on millions of Americans, have been the subject of breaches and the targets of break-ins. Government election offices, political campaigns and data analytic companies working on elections had breaches of datasets containing most or all the personal information an attacker needed to impersonate voters at the 36 websites. Copies of some of these datasets are publicly available for anyone, including an attacker, to access on the deep web (e.g., [65], [51]).

In presenting our detailed findings, we walk through two approaches for an attacker to acquire the personal voter information needed at the 36 state websites; see Figure 2. In the first approach, we start with (a) voter lists, and then append requisite (b) birth and (c) gender information to complete the set of demographics needed per state website (Table 2). We then use the compiled demographic data to acquire any needed government-issued numbers, such as (d) Social Security numbers and (e) driver's license data. In the second approach, the attacker acquires all the data needed from clandestine sources on the (f) darknet or elsewhere on the (g) deep web. Our lists of available sources and any data quality issues, access requirements, or costs appear in the following subsections.

Results for Step 3: Data Sources: (a) Voter Lists

"Voter lists" contain the name, personal demographics, and voting history of each registered voter. As an example, Table 5 list the fields of data available in the South Carolina voter list; the list includes name, date of birth, gender, race, address, and voting history for each registered voter. Government election offices tend to share some (or all) of these fields widely to help assure the integrity of the voting franchise. Listing registered voters by name with demographic information makes the voting process more transparent and election results more accountable. An attacker, seeking to impersonate voters on websites, might first acquire personal information about registered voters.

The local election office is usually responsible for the integrity of the local voting list. Voting lists are almost always available at the local or county level. As described earlier, statewide systems have emerged, and so state election offices control access to statewide voting lists and may also provide a copy of the voting list for a given county. Therefore, there are usually two government sources of voter lists, one at the local or county level and another at the state level.

An attacker who wants to target a specific municipality could acquire the voting list from the local election office or from the state election office. Access requirements and costs often vary by municipality, county, or state for the same data. So, shopping around can make a difference. Of course, if the attacker needs access to the entire state's voter list, then assembling the list using local sources could be onerous, as each locality would have to be contacted separately and each source may have a different process for obtaining a voter list. On the other hand, obtaining voter data from a local election office is usually easier and less restricted than obtaining voter data at the state level. Similarly, if statewide data are expensive, acquiring voter data through local sources may cost less. As an example, we acquired the voter list for the city of Cambridge, Massachusetts for $20, even though the state of Massachusetts would not directly provide the statewide voter list to us.

In this writing, we focus on statewide voter lists for convenience. That is, our investigation surveyed access and costs to obtain statewide voter lists. Acquiring voter data from local municipalities may yield different results.

Access to Voter Lists. Whether a person or organization can acquire a voter list directly from a state election office depends on the state and the stated purposes of the recipient. According to U.S. Voter List Information at the United States Election Project [66] and our own independent survey, state offices for 30 (83 percent) of the 36 websites listed in Figure 4 give or sell their voter lists to anyone in the public, although restrictions exist on acceptable uses. State offices in 23 (64 percent) of the 36 states give or sell their voter lists to anyone in the general public, provided the recipient attests to not using the data for commercial purposes [66]. State offices in 6 of the 36 states, namely Illinois, Kentucky, Maryland, Massachusetts, Minnesota, and South Carolina, do not give or sell their voter list to members of the public but they do give or sell voter lists to specific groups. For example, state election offices in Illinois, Kentucky, Massachusetts, and South Carolina give or sell their voter lists to political campaigns and political committees. State offices in Maryland and Minnesota sell their voter lists to political campaigns and political committees for non-research political purposes. Kentucky and Massachusetts election offices additionally give or sell their voter lists to non-profit organizations. South Carolina's election office does not provide data to the public generally, but it does provide its voter list to any registered South Carolina voter. Table 6 details access restrictions by state.

To summarize access to government sources of voter data: an attacker in 2016 could request voter lists from state offices in 30 (83 percent) of the 36 states that are the subject of this study. In the other 6 states, the attacker would have had two options: collude with a political committee, non-profit organization, or a registered voter for access to the statewide government offering, or request copies of localized lists from municipal or county offices.

County Code	Residential Zip-Code	Date Last Voted	Township
Voter Id	Mailing Number	General Last Voted	House District
First Name	Mailing Name	Primary Last Voted	Senate District
Middle Name	Mailing City	Democrat Last Voted	County Council
Last Name	Mailing State	Republican Last Voted	School Board
Suffix	Mailing Zip	Dem Primary Last Voted	City Council
Residential Number	Gender	Rep Primary Last Voted	Congressional
Residential Name	Race	Watershed District	Status
Residential Suite	Registration date	Precinct	Municipal Code
Residential City	Date of Birth	Precinct Code	Municipal Name

Table 5. Data fields in the South Carolina statewide voter list, includes the Voter ID, name, date of birth, address, gender, race, voting history, voter status, and precinct information for each registered voter.

	Statewide Voter List Access Requirements			
	Any Person or Organization	State Residents	Non-Profit Organizations	Political Committees or Candidates
Alabama	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes

	Statewide Voter List Access Requirements			
	Any Person or Organization	State Residents	Non-Profit Organizations	Political Committees or Candidates
Alaska	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Arizona	Political purposes (and research)	Political purposes (not research)	Political purposes (not research)	Political purposes (not research)
California	Political purposes (not research)	Political purposes (not research)	Political purposes (not research)	Non-commercial purposes
Colorado	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Connecticut	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Delaware	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Election purposes
Georgia	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Hawaii	Election purposes	Election purposes	Election purposes	Non-commercial purposes
Illinois	Cannot use	Cannot use	Cannot use	Non-commercial purposes
Indiana	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Kansas	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Kentucky	Cannot use	Cannot use	Non-commercial purposes	Non-commercial purposes
Louisiana	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Maryland	Cannot use	Cannot use	Cannot use	Political purposes (not research)
Massachusetts	Cannot use	Cannot use	Non-commercial, not researchers	Non-commercial purposes
Michigan	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Minnesota	Cannot use	Cannot use	Cannot use	Political purposes (not research)
Missouri	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Nebraska	Political purposes (not research)	Political purposes (not research)	Political purposes (not research)	Political purposes (not research)
Nevada	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
New Jersey	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
New Mexico	Political purposes (not research)	Political purposes (not research)	Political purposes (not research)	Non-commercial purposes
New York	Election purposes	Election purposes	Election purposes	Election purposes

	Statewide Voter List Access Requirements			
	Any Person or Organization	State Residents	Non-Profit Organizations	Political Committees or Candidates
Ohio	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Oregon	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Pennsylvania	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Rhode Island	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
South Carolina	Cannot use	Non-commercial, registered voters	Non-commercial purposes	Non-commercial purposes
Texas	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Utah	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Vermont	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Virginia	Election purposes	Election purposes	Election purposes (not researchers)	Non-commercial purposes
Washington	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
Washington, DC	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes
West Virginia	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes	Non-commercial purposes

Table 6. Access requirements for statewide voter lists in the 35 states and D.C. from Figure 4. States have requirements that vary by recipient: any person or organization, a state resident, a non-profit organization, or a political committee or political campaign. States also vary access requirements by data uses. Non-commercial use means the sale or resale of the data is prohibited. Political purposes include election activities, political activities, and voter registration. Primary source: U.S. Voter List Information at the United States Election Project [66].

Some voter lists are freely available online for anyone to download. For example, the state of Ohio allows anyone to download the most recent Ohio voter list online at no charge [48]. Some of those who acquire voter lists from government offices subsequently host the data online for public use. For example, websites exist where anyone can download recent copies of statewide voter lists from Colorado [67], Connecticut [68], Delaware [69], Michigan [70], Rhode Island [71], Utah [72], and the District of Columbia [73].

Data brokers, such as Aristotle [49], Catalist [74], L2 [50], e-Merges [75], and Nation Builder [76], specialize in selling voter data (these are called “voter list brokers” in this writing). These

companies offer national databases as well as statewide versions or extracts by county or voting district. Voter list brokers add value to statewide voter lists in a variety of ways. They usually append additional information on each voter, such as more demographic data, lifestyle facts, hobbies, interests, email addresses, and phone numbers. They standardize the data format because each original source of voter data, whether a local municipality or a state election office, has its own format for storing a voter list. For example, the date of birth, April 5, 1945, might be stored as 19450405 in one voter list and as 05-APR-1945 in another. The order of fields and which fields are present also vary. These differences require time-consuming standardization of the selection and arrangement of fields and the format of values within each field in order for computer programs to use the data from multiple sources. Voter data from a voter list broker has a consistent, standard format. Several voter list brokers also offer advanced computing services on the data, such geo-coding, specialized sorts, mapping, and analyses. An attacker can acquire voter information for any municipality in the United States from a voter list broker.

To summarize access to voter data: An attacker in 2016 could acquire statewide voter lists for all 36 states from government offices and/or voting list brokers.

Prices for Voter Lists. First, we surveyed the price of voter lists from government sources. Table 7 reports the total amount per state as well as a per voter cost. Offices in Massachusetts, New York, Ohio, Vermont, and Washington provided the data for free to qualifying groups. The median price was $180 and the average $2,619, with a standard deviation of $7,786. Arizona had the most expensive list at $34,006. All costs were surveyed in 2016.

Next, we surveyed the price of voter lists for the 36 states from 3 voter list brokers. Aristotle offered a rate of 3 cents per voter on its website [49] and sold statewide voter lists at a negotiable rate. L2 had a slightly higher per voter rate of $0.032 on its website [50], and its website in January 2016 offered a flat rate of $2,000 per state for any statewide voter data. The e-Merges website [75] offered a statewide voter list with prices ranging from $350 for D.C. to $10,000 for Arizona, with a median price of $750 and an average price of $2,186.

State		Number of Registered Voters	Authoritative Government Source of Statewide Voter List		
			Direct	Per Voter	Indirect
Alabama	**[77]**	3,330,802	$33,308.00	$0.01000	
Alaska	**[78]**	512,832	$20.00	$0.00040	
Arizona	**[79]**	3,400,628	$34,006.28	$0.01000	
California	**[80]**	17,028,290	$30.00	$0.00000	
Colorado	**[81]**	3,571,127	$50.00	$0.00001	statewide 2016 [67]
Connecticut	**[82]**	2,115,434	$300.00	$0.00014	statewide 2016 [68]
Delaware	**[83]**	681,950	$10.00	$0.00001	statewide 2015 [69]

State		Number of Registered Voters	Authoritative Government Source of Statewide Voter List		
			Direct	Per Voter	Indirect
Georgia	[84]	4,695,698	$250.00	$0.00005	
Hawaii	[85]	706,890	$500.00	$0.00071	
Illinois	[86]	8,045,141	$500.00	$0.00006	
Indiana	[87]	4,829,243	$5,000.00	$0.00104	
Kansas	[88]	1,644,614	$200.00	$0.00012	
Kentucky	[89]	3,105,349	$450.00	$0.00014	
Louisiana	[90]	3,031,808	$5,000	$0.00165	
Maryland	[91]	3,900,090	$128.00	$0.00003	
Massachusetts	[92]	4,271,835	Free to qualifying groups		
Michigan	[93]	7,454,553	$22.50	$0.00000	statewide 2016 [70]
Minnesota	[94]	3,269,260	$46.00	$0.00001	
Missouri	[95]	4,190,936	$100.00	$0.00002	
Nebraska	[96]	1,140,741	$500.00	$0.00044	
Nevada	[97]	1,686,407	$500.00	$0.00030	1 county 2010 [112]
New Jersey	[98]	5,819,276	$2.55	$0.00000	
New Mexico	[99]	1,197,953	$4,791.81	$0.00400	
New York	[100]	12,493,250	Free for election activities		
Ohio	[48]	7,861,025	Free online download		statewide 2016 [48]
Oregon	[101]	2,568,602	$500.00	$0.00019	
Pennsylvania	[102]	8,722,279	$20.00	$0.00000	
Rhode Island	[103]	770,875	$25.00	$0.00000	statewide 2015 [71]
South Carolina	[104]	2,722,287	$160.00	$0.00006	
Texas	[105]	15,101,087	$1,271.95	$0.00008	
Utah	[106]	1,558,418	$1,050.00	$0.00067	statewide 2014[72]
Vermont	[107]	469,833	Free for non-commercial uses		
Virginia	[108]	5,529,742	$5,035.00	$0.00091	
Washington	[109]	4,270,270	Free for non-commercial uses		
Washington, DC	[110]	479,802	$2.00	$0.00000	DC 2016 [73]
West Virginia	[111]	1,274,887	$500.00	$0.00039	

Table 7. Prices for statewide voter lists in the 35 states and D.C. when provided directly from respective state election offices ("authoritative" or "government" source), with the number of registered voters for each state and the per voter cost. Recent copies of government-issued voter lists are freely available indirectly at third-party websites. The most recent years of data available from indirect sources are shown. References appear in []'s.

An attacker could price shop. Suppose an attacker wanted to target all 36 websites. (1) He could purchase statewide data from a voter list broker for $2,000 per state, for a total of $72,000. (2) Alternatively, he could purchase statewide data directly from government sources or indirectly through qualified groups for a total of $94,279, which he could lower to $92,820 by using third-party websites that offer free downloads of recent lists as available. Or (3) he could strategically reduce the cost to $17,679 by downloading freely available voter lists from third-party websites and purchasing lists from government sources for states having prices $2,000 or less and from voter list brokers for all others. Table 8 summarizes costs under all three plans.

Alternatively, an attacker could opportunistically target websites based on available funds. For $219, an attacker could acquire voter data for 18 (50 percent) of the 36 websites from freely available third-party websites and states that charge less than or equal to $100 per list. These 18 states are: Alaska, California, Colorado, Connecticut, Delaware, District of Columbia, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, Utah, Vermont, and Washington. See the rightmost column of Table 8. For $957, an attacker could purchase voter lists from 21 (or 58 percent) of the 36 states; see the rightmost column in Table 8 for voter lists that cost less than or equal to $250 each. Finally, for $4,407, an attacker could acquire voter data for 29 (81 percent) of the 36 websites; see Massachusetts through Utah in the rightmost column of Table 8 for voter lists that cost less than or equal to $500 each.

The Texas website includes a voter identification number, which is a field in the voter data acquired from Texas, so at $1,272, the State of Texas is itself the least expensive provider of its data from named sources and of the Texas voter identification number.

State	Cost for Statewide Voter Lists by Named Sources		
	(a) Authoritative Government	(b) Cheapest of: Free or Government	(c) Cheapest of: Free, Government, or Broker
Massachusetts	$0	$0	$0
New York	$0	$0	$0
Ohio	$0	$0	$0
Vermont	$0	$0	$0
Washington	$0	$0	$0
Washington, DC	$2.00	$0	$0
New Jersey	$2.55	$2.55	$2.55
Delaware	$10.00	$0	$0
Alaska	$20.00	$20.00	$20.00
Pennsylvania	$20.00	$20.00	$20.00
Michigan	$22.50	$0	$0
Rhode Island	$25.00	$0	$0

State	Cost for Statewide Voter Lists by Named Sources		
	(a) Authoritative Government	(b) Cheapest of: Free or Government	(c) Cheapest of: Free, Government, or Broker
California	$30.00	$30.00	$30.00
Minnesota	$46.00	$46.00	$46.00
Colorado	$50.00	$0	$0
Missouri	$100.00	$100.00	$100.00
Maryland	$128.00	$128.00	$128.00
South Carolina	$160.00	$160.00	$160.00
Kansas	$200.00	$200.00	$200.00
Georgia	$250.00	$250.00	$250.00
Connecticut	$300.00	$0	$0
Kentucky	$450.00	$450.00	$450.00
Hawaii	$500.00	$500.00	$500.00
Illinois	$500.00	$500.00	$500.00
Nebraska	$500.00	$500.00	$500.00
Nevada	$500.00	$500.00	$500.00
Oregon	$500.00	$500.00	$500.00
West Virginia	$500.00	$500.00	$500.00
Utah	$1,050.00	$0.00	$0.00
Texas	$1,271.95	$1,271.95	$1,271.95
New Mexico	$4,791.81	$4,791.81	$2,000.00
Indiana	$5,000.00	$5,000.00	$2,000.00
Louisiana	$5,000.00	$5,000.00	$2,000.00
Virginia	$5,035.00	$5,035.00	$2,000.00
Alabama	$33,308.00	$33,308.00	$2,000.00
Arizona	$34,006.28	$34,006.28	$2,000.00
Totals	**$94,279.09**	**$92,819.59**	**$17,678.50**

Table 8. Prices of statewide voter lists in the 35 states and DC from named sources sorted by cost. Column (a) list prices for voter lists directly from state government offices. Column (b) lists "$0" if the voter list could be obtained from free downloads either from the state or from a free website, if available. Otherwise, it lists the price from the state government office. Column (c) lists "0" if the voter list could be obtained through free downloads. If not, it lists the cost from government offices only if government price is less than $2000. If the government price is more than $2000, it lists the price from the voter list brokers. In short, column (c) reflects the least expensive option to buy a voter list from either a free download,

a state government office, or a voter list broker. See Table 7 for more details. The least expensive option is (c) at $17,678.50 for data on all voters in the 35 states and DC.

Results for Step 3: Data Sources: (b) Dates of Birth

We purchased voter lists from government offices and from voter list brokers. We downloaded voter lists from third-party websites. We reviewed descriptions that government offices and voter list brokers provided of their voter lists. Finally, we constructed a national voter database to determine which fields of personal information appear in voter lists.

First, we surveyed the fields of voter lists provided from government sources. Twenty-three (64 percent) states provided names and all the demographic information –namely, date of birth, gender, address– needed to impersonate voters at their websites. These states were: California, Connecticut, Illinois, Kansas, Kentucky, Maryland, Massachusetts, Missouri, Nebraska, Nevada, New Jersey, New Mexico, New York, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Texas, Utah, Virginia, Washington, and West Virginia (see path 3A on Figure 8). Some of these state websites still required additional government issued numbers –namely, driver's license data and Social Security numbers– as described later.

Thirteen states had websites that require an attacker to enter a voter's date of birth, but provided voter lists that do not include dates of birth. These states were: Alaska, Alabama, Arizona, Colorado, Delaware, Georgia, Hawaii, Indiana, Louisiana, Michigan, Minnesota, Vermont, and the District of Columbia (see path 3B on Figure 8). The voter lists from some states, e.g., Alabama and Louisiana, provide age. Other states, e.g., Colorado, Delaware, Michigan, Minnesota, and Vermont, provide year of birth. Path 3 on Figure 8 has a state-wise breakdown.

When we surveyed the fields of data provided by voter list brokers, we found similar results. Voter list brokers provided data for 24 (67 percent) states that contained all the name and demographic information needed to impersonate voters at those state websites. The states were: Alabama, Arizona, California, Connecticut, Illinois, Indiana, Kansas, Kentucky, Maryland, Massachusetts, Nebraska, Nevada, New Jersey, New York, Oregon, Pennsylvania, Rhode Island, South Carolina, Texas, Utah, Virginia, Washington, the District of Columbia, and West Virginia (see path 1A on Figure 8).

Voter list brokers provided data for 11 states that did not have the date of birth information needed at the states' websites. These states were: Alaska, Colorado, Delaware, Georgia, Hawaii, Louisiana, Michigan, Minnesota, New Mexico, Ohio, and Vermont (see path 1B on Figure 8).

We surveyed data brokers to acquire missing date of birth information. There are many kinds of data brokers. Earlier in this writing, we examined data brokers that specialized in voter data and termed them voter list brokers. In this section, we introduce data brokers that

specialize in compiling dossiers from disparate publicly available personal information; we term these "public data brokers" in this writing.

Public data brokers tend to gather and merge personal data from diverse public sources, including some combination of: (1) population registries, e.g., marriage records, divorce records, marriage license records, driving records, phone number directories, professional licenses, weapons permits, and voter registrations; (2) court records, e.g., civil filings, civil actions, bankruptcies, liens, and judgments; (3) property records, e.g., mortgage records and real estate property and owner information; (4) criminal records, e.g., arrests and warrants, mug shots, DUI (driving under the influence of alcohol) records, restraining orders, misdemeanors and felonies, convictions and incarcerations, and driving violations; and, (5) online and social media, e.g., email addresses, and personal references appearing on Facebook, Flickr, Twitter, Google+, MySpace, Kickstarter, LinkedIn, YouTube, Picasa, Pinterest, Reddit, Gravatar, Ancestry, Klout, Instagram, Amazon, Vimeo, and Foursquare.

A simple Google search of a voter's name, city, and state (and optionally "date of birth") revealed links to public data brokers that optimize part of their database content for web searches. We harvested links from search results and ads to compile a list of 15 public data brokers: People Smart (www.peoplesmart.com), BeenVerified (www.beenverified.com), Spokeo (www.spokeo.com), PublicRecords (publicrecords.directory), Radaris (radaris.com), Intelius (www.intelius.com), People by Name (www.peoplebyname.com/people/), Public Data Check (publicdatacheck.com), Truth Finder (www.truthfinder.com), Spyfly (spyfly.com), Instant Checkmate (www.instantcheckmate.com), White Pages (www.whitepages.com), People Finders (www.peoplefinders.com), Persopo (persopo.com), and DOB Search (dobsearch.com). In this writing, we consider these "popular public data brokers".

Freely available search results on 13 (87 percent) of the 15 popular public data brokers show a voter's age when the search contains the voter's name, city, and state. The other two public data brokers provide more detailed birth information. Been Verified sometimes reports the full date of birth and other times reports only age in free searches. People by Name reports the month and year of birth in free searches. Eleven (or 73 percent) of the 15 public data brokers claimed to provide the date of birth for a fee as part of a complete dossier on the person. Prices for one month of unlimited searches ranged from $8.99 (DOB Search) to $34.78 (Instant Checkmate), with an average monthly price of $26.01.

In short, for less than $30, an attacker could search for known voters at popular public data broker websites in order to append date of birth information to voter lists. (See a later section on automation, Results for Step 5. Test automation, for examples of search automation.)

Results for Step 3: Data Sources: (c) Inferring Gender

Gender is another field an attacker needs in order to impersonate voters at the websites for Hawaii, Louisiana, Missouri, New Mexico, South Carolina, and Washington (see Table 3). Voter lists acquired from government sources included gender.

Data from voting list brokers included gender for all the state websites that require it except for Missouri. There are several strategies an attacker could have deployed to enter gender on the Missouri website. One option is for an attacker to enter the same gender for all voters whose address the attacker wants to change. Missouri's system would presumably not change the addresses of those voters for whom the wrong gender was entered but would change the addresses of those voters for whom the correct gender was entered. So, about half the attempted record changes would succeed.

As a second option, an attacker can use lists of common male and female first names (e.g., [113]). When a first name on the voter list agrees with a name most often associated with one gender, the attacker enters the respective gender. The attacker can ignore all voters whose first names are not on the list or whose first name is not easily distinguishable.

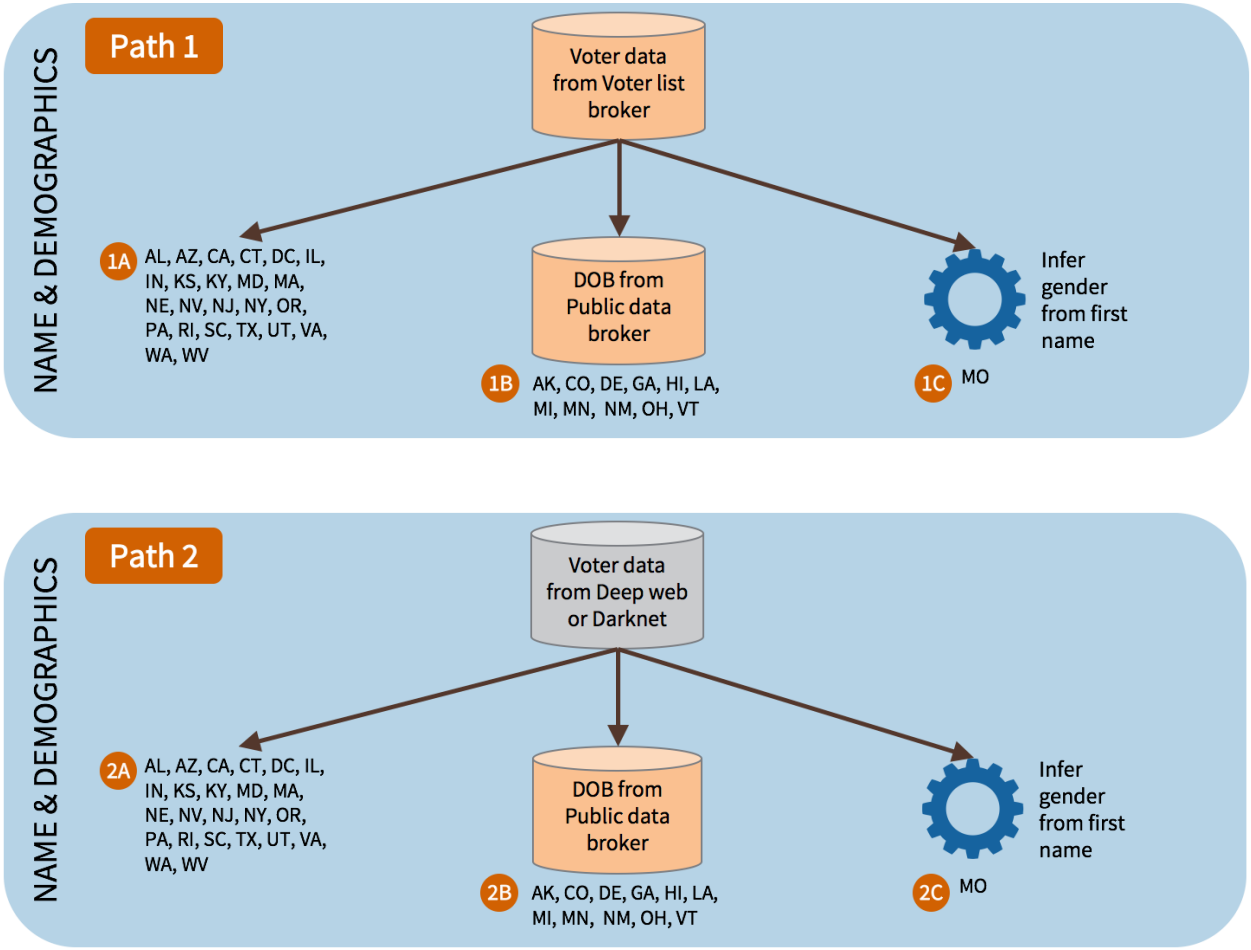
A final option is for an attacker to try one gender, and if it fails for a voter, retry with the other gender. In this case, some attempts to change a voter address would take two tries to be successful. This strategy can also work with eye color. If eye color is a required field, the attacker can try attempt multiple changes until successful since there are a limited number of eye color options. The attacker could pursue any of these three options by manual or automated means.

In short, an attacker could have acquired all the demographic information needed to impersonate a voter on any or all of the 36 websites listed in Figure 4 by obtaining some combination of voter lists and date of birth information from public data brokers, and, for the Missouri website, inferring gender.

Figure 8 summarizes the attacker's data acquisition requirements for each state. An attacker could have acquired statewide voter lists for any of the 35 states and D.C. from either a voting list broker (Path 1) or a government source (Path 3).

Voting lists from a voter list broker provided the names and personal demographics required for the websites of Alabama (AL), Arizona (AZ), California (CA), Connecticut (CT), Illinois (IL), Indiana (IN), Kansas (KS), Kentucky (KY), Maryland (MD), Massachusetts (MA), Nebraska (NE), Nevada (NV), New Jersey (NJ), New York (NY), Oregon (OR), Pennsylvania (PA), Rhode Island (RI), South Carolina (SC), Texas (TX), Utah (UT), Virginia (VA), Washington (WA), District of Columbia (DC), and West Virginia (WV); these are identified on path 1A in Figure 8. No further demographic information was necessary for these states. However, gender was needed for Missouri, and date of birth was needed for the other 11 states. A public data broker could have provided the missing date of birth information, as identified by path 1B on Figure 8. Gender could be inferred from first names (see path 1C on Figure 8).

Similarly, voting lists from government sources provided the names and personal demographics required for the websites of California (CA), Connecticut (CT), Illinois (IL), Kansas (KS), Kentucky (KY), Maryland (MD), Massachusetts (MA), Missouri (MO), Nebraska (NE), Nevada (NV), New Jersey (NJ), New Mexico (NM), New York (NY), Oregon (OR), Pennsylvania (PA), Rhode Island (RI), South Carolina (SC), Texas (TX), Utah (UT), Virginia (VA), Washington (WA), and West Virginia (WV), as identified by path 3A on Figure 8. No further demographic information was necessary for these states. However, date of birth was needed for the remaining 12 states. A public data broker could provide the missing date of birth information, as identified by path 3B on Figure 8.



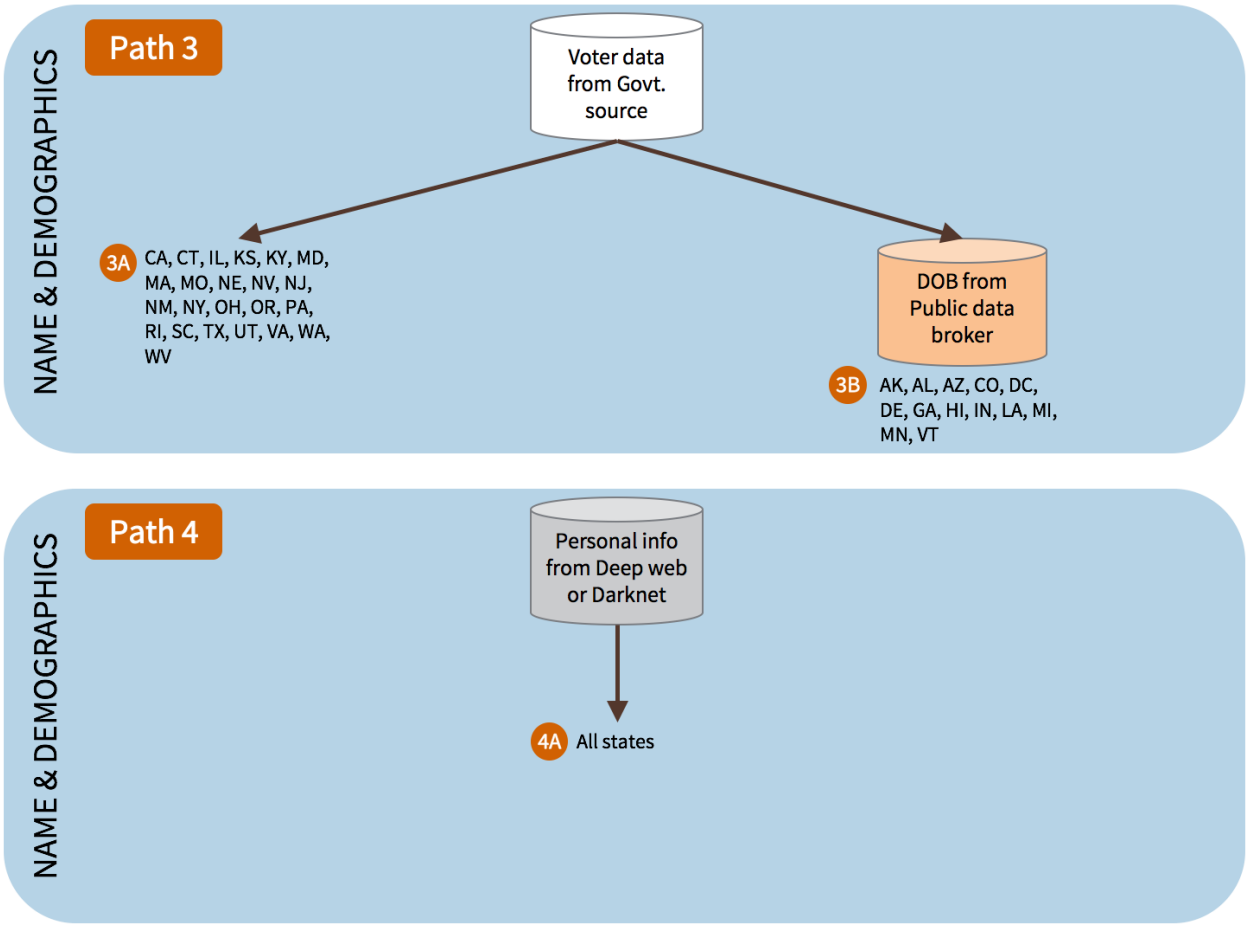
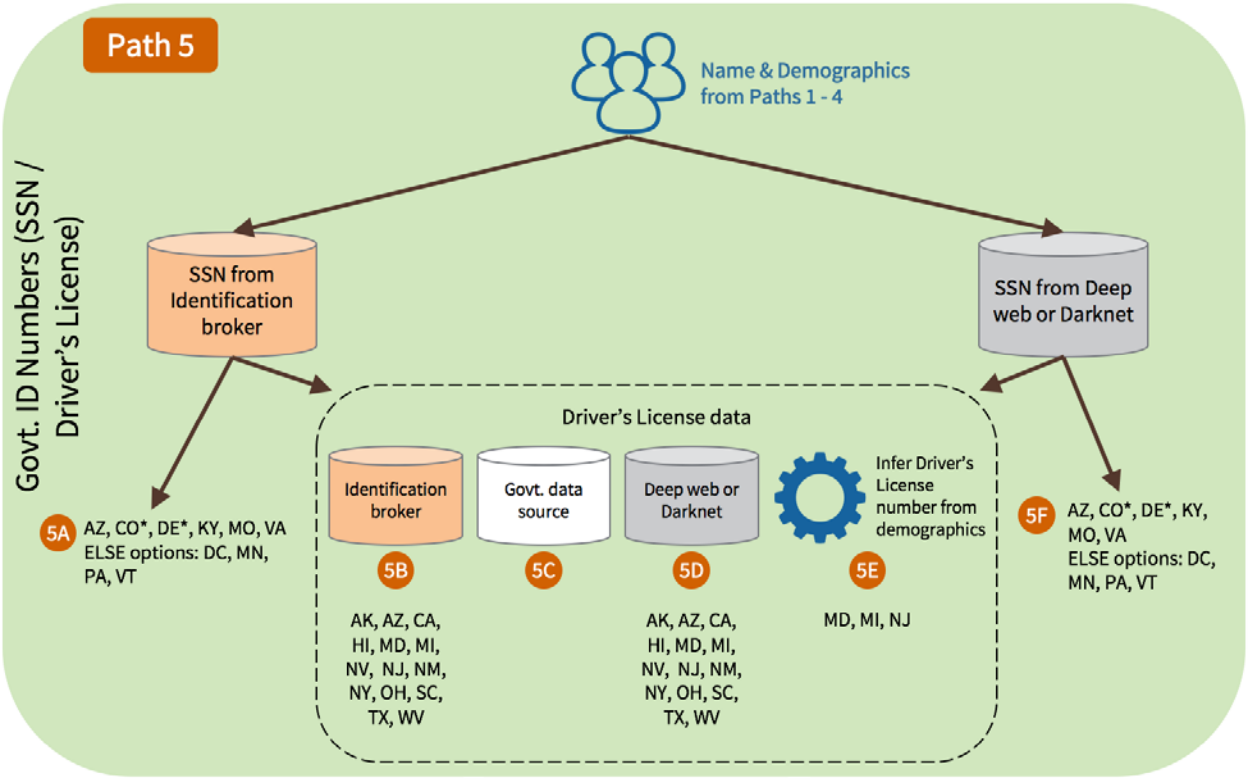
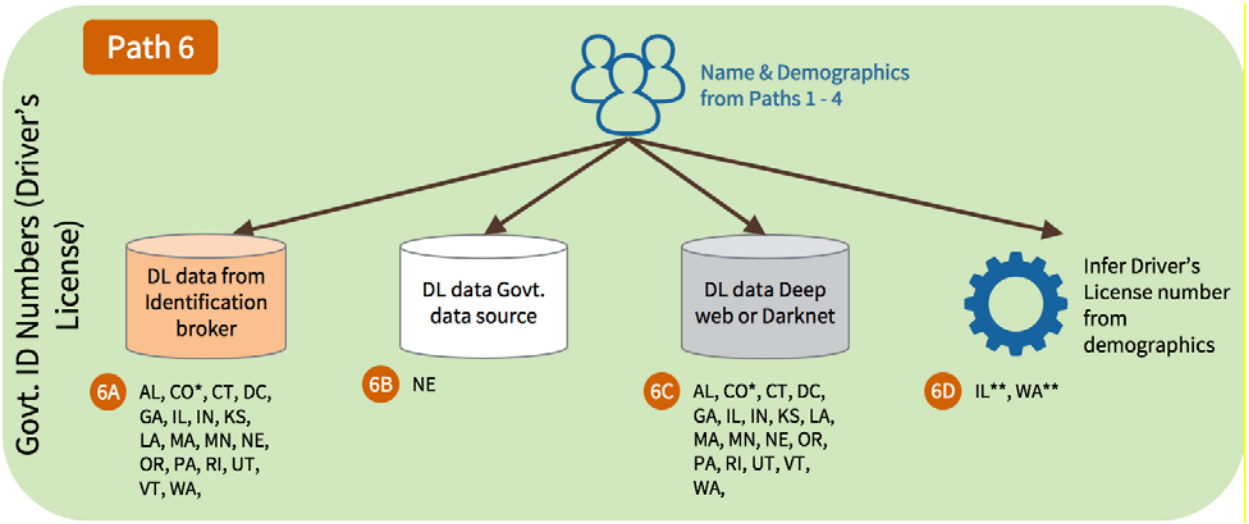


Figure 8. Four paths for an attacker to acquire the names and demographics needed to impersonate voters on the 36 state websites listed in Figure 4. Paths 1, 2, and 4 follow data from a voter list broker or darknet or deep web source. Path 3 follows data from government sources. Path 1A identifies states (abbreviations AL to WV) for which acquired voter lists from a broker was sufficient in 2016. Path 1B identifies states (abbreviations AK to VT) for which voting lists from brokers augmented with date of birth information were sufficient. Path 1C specifies the need to add gender to a Missouri data (MO) voter list acquired from a broker. Path 3A identifies states (abbreviations CA to WV) for which acquired voter lists from government sources were sufficient. Path 3B identifies states (AK to VT) for which voting lists from government offices augmented with date of birth information were sufficient. Alternatively, names and demographics are available with SSNs or driver's license data on the darknet or deep web as shown by Path 4A. Paths continue on Figure 9, which shows ways to append any needed government identifiers.


Path 5
Name & Demographics from Paths 1 - 4
Govt. ID Numbers (SSN / Driver's License)
SSN from Identification broker
SSN from Deep web or Darknet
Driver's License data
Identification broker
Govt. data source
Deep web or Darknet
Infer Driver's License number from demographics
5A AZ, CO*, DE*, KY, MO, VA ELSE options: DC, MN, PA, VT
5B
5C
5D
5E
AK, AZ, CA, HI, MD, MI, NV, NJ, NM, NY, OH, SC, TX, WV
AK, AZ, CA, HI, MD, MI, NV, NJ, NM, NY, OH, SC, TX, WV
MD, MI, NJ
5F AZ, CO*, DE*, KY, MO, VA ELSE options: DC, MN, PA, VT


Path 6
Name & Demographics from Paths 1 - 4
Govt. ID Numbers (Driver's License)
DL data from Identification broker
DL data Govt. data source
DL data Deep web or Darknet
Infer Driver's License number from demographics
6A AL, CO*, CT, DC, GA, IL, IN, KS, LA, MA, MN, NE, OR, PA, RI, UT, VT, WA,
6B NE
6C AL, CO*, CT, DC, GA, IL, IN, KS, LA, MA, MN, NE, OR, PA, RI, UT, VT, WA,
6D IL**, WA**

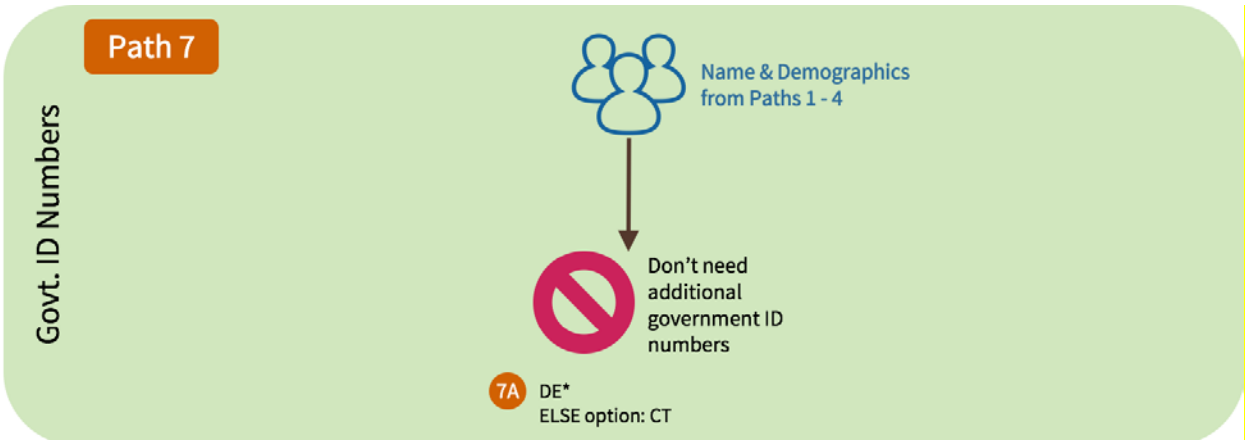


Figure 9. Continuation of paths 1 through 4 from Figure 8 to additionally acquire any SSNs or driver's license data needed to impersonate voters on the 36 state websites listed in Figure 4. In Figure 8, Paths 1 through 4 show how names and demographic information might be acquired from a voter list broker, government source, or the deep web or darknet. Path 5 applies to states that need driver's license and SSNs and Path 6 to states that need driver's license only. As shown in Path 7A, two states required no additional information. Path 5A concludes with SSNs acquired from an identification broker for 6 states (AZ to VT) and the ELSE requirement for 4 states (DC to VT). Path 5A provides SSNs for the same states as does path 5F, except that the SSNs come from sources on the deep web or darknet. Driver's license data is the driver's license number and issue date. Path 5B and 6A conclude the requirements for 32 states (AK to WV) with driver's license data from an identification broker. Path 6B satisfies NE requirements using a government website. Path 5E and 6D satisfy requirements for 5 states using an algorithm. Path 5D and 6C provide driver's license data for the same states as does path 5B and 6A, except that the data come from sources on the deep web or darknet. * indicates an OR requirement (Table 2); ** states also need the driver's license issue date, which is not included in the algorithmic computation.

Results for Step 3: Data Sources: (d) Social Security Numbers

None of the voter lists acquired from government sources or from voter list brokers included Social Security numbers. We surveyed other kinds of data brokers as a means for an attacker to obtain the Social Security numbers of named voters needed on the 17 (47 percent) websites that require SSNs.

Voter list brokers, as described earlier, provide voting lists and related services. One voter list broker may be better than another based on how recent or complete the data may be and added fields of information. Public data brokers, as described earlier, gather public records and produce dossiers on individuals primarily from publicly available information. One public data broker may be better than another based on sources of information and the ability to correctly link information about the same person across disparate public sources. Most sources of public records merely record the person's name and address; some include the

person's date of birth. Therefore, attempts to construct personal dossiers from public records necessitate matches of name, address and date of birth. False and missed matches are common.

One way to dramatically improve linkages is to incorporate sources that record Social Security numbers and driver's license numbers. Often these are financial, insurance, employment, or credit-related sources of personal information. In this writing, we use the term "identification broker" to refer to a data broker that provides personal information that includes Social Security numbers or driver's license numbers. Of course, the fact that a data broker has or uses Social Security numbers and driver's license numbers internally does not necessarily mean that the broker provides or sells those numbers to others. We reserve the term "identification broker" for a broker that provides either Social Security numbers, driver's license numbers, or both for named individuals. Identification brokers often provide services related to verifying identity, such as verifying information on rental applications or performing background employment checks.

Any particular data broker may offer a suite of data options; therefore, a broker may be a voter list broker, a public data broker, and an identification broker, or some combination of these.

In order to locate identification brokers, we started with previously assembled lists. In prior work, Pam Dixon at the World Privacy Forum [114] and Julia Angwin at Propublica [115] itemized data brokers who held information on millions of Americans as part of their own separate studies. The World Privacy Forum listed 347 brokers [114], and Propublica listed 201 sources [115]. We combined these lists and acquired a list of 426 distinct brokers, to which we added 5 popular data brokers from above that were on neither list. We also added 23 other data brokers we learned about during our investigations. Our list began with 454 brokers.

We surveyed the websites of each of the 454 brokers and recorded any claims made to provide dates of birth, Social Security numbers or driver's license numbers for Americans, and the price for doing so. We then made purchase agreements with 5 brokers to verify available content.

Thirty-five of the 454 data broker websites were defunct, and 20 websites no longer worked. That reduced the total number of operational data broker websites to 399.

The websites of 79 data brokers offered Social Security number services, and the websites of 49 data brokers provided driver's license number services. Fifty additional data brokers, beyond those that provide SSNs or driver's license numbers, advertised an ability to provide date of birth information on named Americans.

Of the 79 data broker websites offering Social Security number services, 14 offered the SSNs of decedents only. Two data broker websites reported whether a given Social Security number had been assigned and if so, whether the recipient was deceased. Thirty-seven (37)

data broker websites verified matches between Social Security numbers and names. Twenty-two (22) websites provided the ability to associate a name with a given Social Security number or a Social Security number with a given name and address or birthdate.

Of course, the last group, the 22 broker websites that provide an SSN in response to a search of a voter's name and address or to a search additionally including the voter's date of birth or age, would be the most convenient service for an attacker. Prices varied widely, making it cost prohibitive to try each of the 22 data brokers. However, we tested 5 broker websites and found SSNs correctly associated with the searched names in all cases. (See a later section on automation, Results for Step 5. Test automation, for examples of search automation.) Our maximum per search cost was $1 and our minimum cost was an unlimited number of searches (seemingly for a single machine) for a monthly fee of $39.95.

The Gramm-Leach-Bliley Act (GLBA) is a federal law enacted in the United States to control the ways that financial institutions handle the personal information of individuals [116]. It covers the kind of personal information, including Social Security numbers, that individuals provide to open a checking account, acquire an insurance policy, apply for a loan, or request a line of credit. If a data broker uses information from a GLBA source, then any user who accesses the data broker's data must attest to using the information for a "permissible purpose." Permissible purposes under the GLBA are: (a) to protect against or prevent actual or potential fraud, unauthorized transactions, claims, or other liability; (b) required institutional risk control, or for resolving consumer disputes or inquiries; (c) use by persons holding a legal or beneficial interest relating to the consumer; or, (d) to comply with Federal, State, or local laws, rules, and other applicable legal requirements. Most data broker websites list the permissible purposes and request website users to mouse-click on the purpose that applies. Self-attestation is sufficient with no further validation. As researchers in this study, we selected (a), as would an individual looking up their own information. Of course, "disenfranchising voters" is not listed; an attacker would have to claim one of these purposes in order to proceed at sites requiring a GLBA selection.

Acquiring SSNs, when combined with the acquisition of demographics we described earlier, completes the requirements for Colorado, Delaware, Kentucky, Missouri, and Virginia; see Table 3. If an attacker wanted to focus on voters who were unlikely drivers, such as the elderly or those whose address was a care institution, then acquiring SSNs would also complete the requirements for Minnesota, Pennsylvania, Vermont and the District of Columbia. These are summarized on Figure 9 as path 5A.

For completeness, we now discuss ways some of the other SSN services could have helped an attacker to impersonate voters, and we identify which other SSN services would not have been beneficial to an attacker. A service based solely on decedent information is not useful because the attacker would only be interested in live voters. On the other hand, searches by SSNs could have helped the attacker in cases where the service permitted an unlimited number of searches at a reasonable price. SSNs assigned prior to 2011 follow an encoded

format that allows predictions about the first 5 digits of the 9-digit number [117]. For voters born after 1987, more of the 9 digits may be predicted from birthdate and birthplace [118]. With some work, an attacker may predict digits of a voter's SSN by assuming the voter was born in the same state. The attacker would then exhaustively search numbers having the same first 3 or 5 digits on the SSN website to determine which of these SSNs are assigned to living people. Then, the attacker would search those numbers assigned to living people to see which match a given voter's name. This approach could be automated to scale performance (see section Results for Step 5. Test automation for examples). However, the direct search approach, described in the prior paragraphs, provides more SSN-to-voter matches with far less effort.

Results for Step 3: Data Sources: (e) Driver's License Data

Recall, 33 (92 percent) of the 36 websites require driver's license data. We surveyed: (1) the use of SSNs as driver's license numbers; (2) state websites that revealed driver's license numbers; (3) data brokers that provided driver's license data; and (4) encodings of personal demographics to construct driver's license numbers. We summarize per state findings in Figure 9.

Several sources (e.g., [119]) have reported that some states still use a person's SSN as a driver's license number. These reports cite Arizona, Indiana, Kansas, Massachusetts, Missouri, Nevada, New Mexico, Ohio, and Virginia. We tested each and only found SSNs somewhat used as driver's license numbers in Arizona, so we appended Arizona to path 5A in Figure 9 for cases because the attacker would only need SSNs.

We surveyed state websites that allowed an individual to inquire about her own driving record or change the address on her driver's license. Our search strings were "<state> driver's license status," "<state> driver's license verify," "<state> driver's license check," and "<state> driver's license record," where <state> is the name of the state that is the subject of the query. We found such websites for 34 of the 36 states; Hawaii and Vermont were the exceptions. We then examined whether these websites reveal a person's driver's license number after one enters the person's name and demographics. Of the 34 websites tested, only the website for Nebraska revealed the driver's license number if an attacker entered a person's SSN [120]. There was no cost for the service. This is summarized on Figure 9 as path 6B.

Of the 454 data broker websites we surveyed for SSN services, 49 offered driver's license services. But only 11 of these provided a driver's license number and issuance date in response to a search of a voter's name and address or date of birth. Monthly subscriptions ranged from $39.95 to $59.95 per month, and some websites charged an additional $1, $5 or $10 per lookup. The $39.95 service provided unlimited searches (seemingly for a single machine) that included SSNs and driver's license numbers and charged $0.01 per record for further details, such as driver's license issue date. We sampled the least expensive of these

and found accurate information. (For completeness, we report that the other 38 data broker websites provided driving reports that listed drunk driving convictions or other driving record information in response to a given driving license number.)

Combining driver's license data from an identification broker with the acquired demographic information described earlier would have provided an attacker with all the data needed to impersonate voters on the websites of Alabama, Colorado, Connecticut, Delaware, Georgia, Illinois, Indiana, Kansas, Louisiana, Massachusetts, Minnesota, Nebraska, Oregon, Pennsylvania, Rhode Island, Utah, Vermont, Washington, and the District of Columbia. These are summarized on Figure 9 along path 6C.

Similarly, when further combined with the SSN information described in the prior subsection, an attacker would have acquired all the data needed to impersonate voters on the websites for Alaska, Arizona, California, Hawaii, Maryland, Michigan, Nevada, New Jersey, New Mexico, New York, Ohio, South Carolina, Texas and West Virginia. These are summarized on Figure 9 along path 5D.

A driver's license number in some states is an encoding of a person's name and date of birth [46], [47]. The relevant states are: Illinois, Maryland, Michigan, Minnesota (licenses issued prior to 2005) Nevada (prior to 1998), New Jersey (prior to 2004), New York (prior to 1992), and Washington. We used information on a random sample of 10 drivers from each state to see whether the encodings remained in use.

Driver's license number encodings often rely on Soundex to phonetically encode a name into a 4-character code [46]. The Soundex code for a name consists of a letter followed by three digits [121]. The letter is the first character of the person's name, and the digits encode the remaining consonants. Similar-sounding consonants share the same digit. Figure 10 shows the Soundex algorithm. Various programs exist on the Web to produce Soundex encodings (such as [122]; some online programs are specific to producing driver's license numbers that use Soundex, such as [47]).

1	Capitalize all letters in the word and drop all punctuation marks.
2	Retain the first letter of the word.
3	Change all occurrence of the following letters to '0' (zero): 'A', E', 'I', 'O', 'U', 'H', 'W', 'Y'.
4	Change letters from the following sets into the digit given: 1 = 'B', 'F', 'P', 'V' 2 = 'C', 'G', 'J', 'K', 'Q', 'S', 'X', 'Z' 3 = 'D','T' 4 = 'L' 5 = 'M','N' 6 = 'R'
5	Remove all zeros from the string.
6	Replace duplicate adjacent values with one instance.
7	Pad the string with trailing zeros and return only the first four positions, which will be of the form <uppercase letter> <digit> <digit> <digit>.

Figure 10. Soundex algorithm for phonetically encoding names to a 4-character code used in some driver's license numbers [46]. "Robert" and "Rupert" are both R163, and "Ashcroft" is A261.

An Illinois driver's license number is based on the driver's name and birth date in the format SSSS-FFFY-YDDD, where SSSS is the Soundex code of the driver's last name, FFF requires looking up the first name and middle initial in a table (see Appendix A [123]), Y-Y is the last two digits of the year of birth, and DDD is the calculation: (birth_month - 1) * 31 + birth_day + (male:0, female: 600). Online programs can compute the number automatically (e.g., [124]). All of the driver's licenses in our sample matched the encoded format. Therefore, an attacker could produce Illinois driver's license numbers from demographics, so we recorded Illinois on path 6D of Figure 9.

Maryland, Michigan, and Minnesota (in licenses issued prior to 2005) use the same driver's license number encoding of the driver's name and birth date in the format LLLL-FFF-MMM-BBB, where LLL is the Soundex code of the driver's last name, FFF is an encoding of the driver's first name based on a table, MMM encodes the driver's middle name based on a table, and BBB is the driver's birthday and month based on a table (see Appendix A [125]). Then, for Michigan, drop the last two digits. Online programs can compute the number automatically (e.g., [126] for Maryland, [127] for Michigan, and [128] for Minnesota). All of the driver's licenses in our sample matched the encoded format for Maryland and Michigan, but not for Minnesota. Therefore, an attacker could produce Maryland and Michigan driver's license numbers from demographics, so we recorded Maryland and Michigan on path 5E of Figure 9.

If issued before 1998, the Nevada driver's license number was based on the driver's SSN and year of birth—specifically, take the SSN, multiply by 2, and then add 2,600,000,001 [129]. Append the two-digit year to the end. The result is a 12-digit value. Nevada driver's licenses issued after 1998 have a two-digit code identifying the issuing office followed by 8 digits. We found none of the older encoded values in our sample.

If issued before 2004, the New Jersey driver's license number is based on the driver's name, birth date, gender, and eye color [130]. The format is Axxxx fffmm MMyye, where A is the first letter of the driver's last name, xxxx is the Soundex encoding of the driver's last name starting with the second letter, fff is the 3-digit part of the Soundex encoding of the driver's first name, mm is an index for the driver's middle initial (A=61, …, Z=89, none=00), MM is the month of birth and gender (01=Jan male, …, 12=Dec male, 51=Jan female, …, 62=Dec female), yy is the last 2 digits of the year of birth, and e is eye color (1=Black, 2=Brown, 3=Grey, 4=Blue, 5=Hazel, 6=Green). All of the driver's licenses in our sample matched the encoded format with guesses for eye color. Therefore, an attacker could produce New Jersey driver's license numbers from demographics, so we recorded New Jersey on path 5E of Figure 9. Of course, guessing eye color would require more attempts.

If issued before 1992, the New York driver's license number is based on the driver's name and birthdate [131]. The first character is the first letter of the last name, and the next 12 digits encode the first three letters in the first name (F1, F2, F2), the middle initial (M1), and the second through fourth letters of the last name (L2, L3, L4 and L5). Each letter has a value (1=A, …, 26=Z, 0 otherwise). The length of the last name is X. Compute: 10,017,758,323 L2 + 371,538,441 L3 + 13,779,585 L4 + 510,355 L5 + 19,657 F1 + 729 F2 + 27 F3 + M1 – X. If the result is not 12 digits long, add zeros to the left side as needed. Then, the next three digits use the driver's birth month (m=1,..,12), birth day (d=1,…,31) and gender (g-0=male, 1=female) to compute: 63 m + 2 d + g. Pad the left side with zeros to make the number three digits long as needed. The last two digits are the driver's year of birth. Finally, if more than one person would have the same number, another digit is added before the year of birth digits, seemingly assigned sequentially. Online programs can compute the number automatically (e.g., [132]). We found none of the older encoded values for New York in our sample.

The Washington driver's license number is based on the driver's name and birth date in the format LLLLLFMYYXmd, where LLLLL is the first five letters of the driver's last name padded with asterisks (*) as needed, F is the initial of the driver's first name, M is the initial of the driver's middle name, YY is 100 minus the two digit year of birth, X is a checksum computed over values from an assignment of codes to prior computed characters (see Appendix A), m is the result of looking up the driver's month of birth in a table of two possible values depending on whether a prior similar driver number has been assigned, and d is the result of looking up the driver's day of birth in a table (see Appendix A [133]). Online programs can compute the number automatically (e.g., [134]). All of the driver's licenses in our sample matched the encoded format. Therefore, an attacker could produce Washington driver's license numbers from demographics, so we recorded Washington on path 6D of Figure 9.

In short, driver's license numbers for Illinois, Maryland, Michigan, New Jersey and Washington could be computed from the voter demographics.

We have now shown multiple means by which an attacker could acquire all the personal data needed to impersonate voters at any of the 36 websites (Figure 4). Figure 8 summarizes the acquisition of needed demographics by state, and Figure 9 describes how to append Social Security numbers and driver's license data as needed. All data sources used so far are named government offices, online websites, algorithms, and data brokers. In the next subsection, we examine how versions of the same data can be acquired through unknown sources on the deep web and darknet.

Results for Step 3: Data Sources: (f) Darknet

In November 2016, February 2017, June 2017, and July 2017, we surveyed darknet websites for offers that included the combinations of personal information that an attacker would need. We found the darknet to be an easily accessible, worldwide marketplace with numerous markets in which personal data on Americans was for sale. Buyers and sellers use

anonymous, but persistent, identities with ratings based on purchase history. Payments are made using bitcoins, often escrowed by the market as a way to assure the integrity of purchases.

Our goal was to learn what kinds of darknet offers might have been available to a 2016 attacker. We did not seek to record the prevalence of personal information on darknet markets. That is, we were not interested in how many offers were available, but in what kinds of offers were available.

Our guidelines were simple. If an offer differed from all other offers we recorded by either the combination of personal information provided, the price, or the geography covered, we included that offer in our count. On the other hand, distinct offers for the same kind of personal information bundled in a similar manner with similar costs and for the same US geography, even if from different suppliers, were counted once.

We used a 2-step approach: (1) use a darknet search engine to identify two markets having the most offers of personal information of interest, and then (2) survey relevant offers on those markets specifically. We did not purchase any data from darknet sources, so actual dataset contents may vary from advertised descriptions. Instead, we use darknet market trust levels (e.g. [135]) as a measure of the likelihood that an offer properly describes the dataset contents. These levels use the reputation of the vendor and range from 1 (no real historical experience) to 10 (having 900 or more sales with 90 percent positive feedback). We accepted offers of level 4 or higher; these are vendors having at least 300 sales with $10,000 volume and 90 percent positive feedback.

First, we searched Grams [136], a search engine used to locate goods for sale on darknet markets. Our search strings were “voter,” “SSN,” “driver’s license,” and “fullz.” The term "fullz" usually contains an individual's name, SSN, birth date, account numbers, and other data [137]. We counted the occurrence of darknet marketplaces on the first 6 pages of search results. The two darknet markets having the most offers from these searches were Alphabay [138] and Hansa [139]. During the writing of this paper, Alphabay was shut down [140], so these offers now appear on Hansa and Dream [141]. Our findings are from our prior surveys of Alphabay and Hansa.

On Alphabay, we recorded the price and contents of the first 25 distinct kinds of offers encountered for access to: (a) datasets that included voter lists or the names, addresses, and dates of birth of thousands of Americans but excluded SSNs or driver’s license data; (b) datasets that also included Social Security numbers; or (c) datasets that included driver’s license data. We then searched Alphabay for the existence of state-specific datasets having the names, addresses, dates of birth, and SSNs of thousands of residents in each state. On Hansa, we surveyed the first 20 offers of (a), (b), or (c) and recorded any that were different in kind from those previously found on the first market. Hansa included links to a master list of datasets called dbworld [142], which seemed to be another means to access offers on Hansa,

so we only counted these kinds of offers once. Viewing the offers on Hansa however, includes a sample of the first rows of the data allowing us to verify contents.

Among what we found for sale were: lists of personal information needed to establish credit, including SSNs and sometimes driver's license data (20 kinds of offers); files of driver's license data (9 kinds of offers); links to online repositories of profiles containing names, demographics, SSNs, and driver's license data (6 kinds of offers); bundles of databases containing names and demographics including voter lists (36 kinds of offers); credit card information including driver's license data (2 kinds of offers); and discounted access to data broker websites often using pre-established account names and passwords (2 kinds of offers). We also found one recurring option for each of the 50 states and D.C. for access to the names, addresses, dates of birth, and SSNs for thousands of that state's residents (51 offers). In total, we archived 95 kinds of offers.

Datasets sold on darknet markets are not necessary illegal. There are many possible legal, illegal, and "in-between" possibilities. Credit card numbers likely originated from illegal activity. Account names and passwords to data broker websites may come from stolen account holder information; or, they may come from account holders exploiting their own pre-existing unlimited-use account arrangements. Opportunists who located breached information freely available online may be the source of lists of names and demographics (including voter lists, or data holders wanting to make money from data they legally hold may be the sources. There are many possibilities.

Regardless of the source of the data, however, pricing structures suggest that primary buyers intend to perpetrate credit card fraud or tax identity theft, because higher prices correlate with increased potential for economic gain through fraud. For example, the most recently released credit card numbers had a higher price than credit card numbers on the market for a longer time, presumably because over time credit card holders and banks cancel the accounts of exposed numbers, rendering the data less useful for fraud. As another example, the SSNs of individuals with higher credit ratings cost more than SSNs of those with lower credit ratings or unknown credit ratings.

These market incentives dictate how an attacker, seeking to impersonate voters on the 36 websites, could acquire the personal information needed. In the prior subsections, we described the attacker's possible trail of data acquisition from named sources as one that started with voter lists, to which dates of birth were added, gender was inferred, and SSNs and driver's license information were acquired as needed (Figure 8 and Figure 9). The shortest trail started with voter lists and then relied on the same data broker to append date of birth, SSNs, and driver's license data.

Using darknet markets, an attacker could acquire access to single datasets having all the fields of information needed. For example, datasets having the personal information needed to establish credit minimally include: name, address, date of birth, and SSN (e.g., the recent

examples in Figure 11 and Figure 12). As mentioned earlier, prices often depend on credit worthiness, so an attacker not concerned with credit could acquire the least expensive options. Prices for these datasets ranged from $0.41 to $3 per SSN record with a median price of $0.90 and negotiable volume discounts. We also found lists of names, addresses and dates of birth and SSNs for all 50 states and the District of Columbia, though not all lists were complete (median price of $0.90 per account). Acquiring any of these datasets for relevant states would have completed the requirements for Colorado, Delaware, Kentucky, Missouri, and Virginia; see Table 3. If the attacker wanted to focus on voters who were unlikely drivers, then acquiring these would have also completed the requirements for Connecticut, Minnesota, Pennsylvania, Vermont, and the District of Columbia. These are summarized in Figure 9 as paths 5F and path 7A. Of course, we do not believe that any of these datasets include all or most adults in a given state, so these datasets would provide information for opportunistic rather than broad attacks.

Offers having the same fields but including most or all of the adult American population are also available. For about $500, an attacker could purchase either a copy of a dataset or obtain unlimited access to a dataset that reportedly contains names and SSNs for all adult Americans who have SSNs (see the recent example in Figure 13). For another $500 (actual price $501.94), an attacker could purchase a dataset having the names, address, gender, and dates of birth of 203 million Americans, seemingly originating from a breach at Experian [30]; see Figure 14. Together, these two datasets would complete the requirements for Colorado, Delaware, Kentucky, Missouri, and Virginia; see Table 3. If the attacker wanted to focus on voters who were unlikely drivers, then acquiring these data also would complete the requirements for Connecticut, Minnesota, Pennsylvania, Vermont and the District of Columbia. These are summarized on Figure 9 as path 5F and path 7A. These datasets would provide comprehensive coverage of the American population.

An attacker could alternatively acquire files of driver's license numbers that include driver's license issue dates, dates of birth, names, and addresses, but not SSNs (e.g., Figure 15). Prices for these datasets ranged from $0.41 to $30 per record with negotiable volume discounts. Again, price depended on credit worthiness. Acquiring one of these datasets would complete the requirements for Alabama, Delaware, Colorado, Connecticut, Georgia, Illinois, Indiana, Kansas, Louisiana, Massachusetts, Minnesota, Nebraska, Oregon, Pennsylvania, Rhode Island, Utah, Vermont, Washington, and the District of Columbia; see Figure 9, paths 7A and 6C.

Because SSNs are critical to credit, their inclusion in a dataset usually has a dominant influence on price. An attacker could just as easily purchase a dataset having SSNs, names, address, dates of birth, and driver's license data as a dataset having SSNs, names, address, dates of birth, and no driver's license data (e.g., Figure 16). Prices of records having SSNs with driver's license data included ranged from $0.41 to $3 with negotiable volume discounts. Acquiring one of these datasets, with SSNs and driver's license data, would complete the requirements for Alaska, Alabama, Arizona, California, Delaware, Colorado, Connecticut,

Georgia, Hawaii, Illinois, Indiana, Kansas, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Nebraska, Nevada, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Texas, Utah, Vermont, Washington, West Virginia and the District of Columbia; see Figure 9 paths 5D, 6C and 7A. These datasets may not include all or most adults in the state.

Of course, the datasets we just described, whether comprehensive or not, included voters and non-voters (those not registered to vote), indiscriminately. An attacker could use these datasets alone for opportunistic attacks, but if an attacker wanted to tamper with an election, a voter list would be useful to identify specific people to impersonate.

Voter lists on the darknet are often difficult to discern from generally available datasets that contain names and addresses of Americans because the seller tends to barely or incidentally mention voter affiliation information among the available fields. For example, Figure 17 shows an example of a bundle of 141 datasets we found for $653, which included voter files for 13 states; an attacker who purchased these data only for the voter files would have paid $13 per million voters.

Voter lists are also offered separately. We found offers for 9 (25 percent of the 36) statewide voter lists; these were: $10.98 for Alaska, $30.97 for Colorado, $25.97 Connecticut, $20.98 Delaware, $30.97 Michigan, $26.97 Nevada, $30.97 Ohio, $25.98 Rhode Island, and $30.97 for Washington. We found offers for 4 voter lists that did not include all the voters in the state; these were: $10.98 for 4 percent of the statewide list from Alabama, $30.98 for 7 percent of the statewide list from Pennsylvania, $15.98 for 4 percent of the statewide list from Texas, and $25.98 for 5 percent of the statewide list from Utah. We also found offers for voter lists from Florida, North Carolina, and Oklahoma, which are not among the 36 states. See Figure 18 for excerpts of these offers. The Texas voter list included the voter ID number for each person.

An attacker could have spent $235 for the 9 statewide voter lists offered over the darknet (Figure 18), instead of spending $678 for those voter lists from government sources (Table 8). The bulk of the savings comes from paying $26.97 for Nevada instead of $500 each to acquire these from named government offices or data brokers. Surprisingly, an attacker could lower the cost further, to just $38, by using freely available online voter lists for Colorado (saving $30.97), Connecticut (saving $25.97), Delaware (saving $20.98), Michigan (saving $30.97), Ohio (saving $30.97), Rhode Island (saving $25.98), and Washington (saving $30.97); see Table 8. Using the least expensive combination of named or darknet sources to acquire statewide voter lists for all 36 states drops the total cost from $17,679 to $17,196.

Also sold on the darknet is discounted access to named data broker websites. Rather than purchasing datasets on the darknet, an attacker could purchase a discount coupon for a named data broker or credentials for a previously established account at a data broker's website. We found prices ranging from $2 to $12, with a median price of $10 per account. The

account credentials (username and password) allow an attacker unlimited use of the data broker website for the flat fee. Figure 19 shows an offer for Intelius, one of the popular public data brokers mentioned earlier. Figure 20 shows offers to URL links where data can be searched freely or downloaded, one of which reports having profiles for 70 percent of all Americans who have SSNs.

Some of the flat fee prices for URL links to personal data on Americans may be URLs hosting files stored freely and openly on the deep web (see the recent examples in Figure 20). Because the URL is freely and openly available, the attacker might be able to locate the data without a darknet purchase at all.

Many of the datasets mentioned may or may not include gender. In cases where the dataset does not include gender, the attacker could infer gender from first names, as described in an earlier section.

The websites for New York and Texas both require additional information from the driver's license beyond the license number and issuance date. New York's website requires a driver's license "document number," and Texas's requires an "audit number" from the driver's license. We refer to both as document numbers, and they are commonly found on U.S. driver's licenses. A driver's license number is unique to a driver but a document number is unique to the card. A replacement driver's license, for example, will have the same driver's license number but a different document number. Different states assign document numbers differently. In Massachusetts, for example, the document number is an encoding of the date of issuances for the driver's license number. Some states may sequentially assign document numbers, so that numbers correlate with issuance dates. Other states may encode other kinds of information, so having a list of these document numbers with other personal information about the drivers will often reveal the encoding scheme. A document number may end with a checksum digit so that a mathematical computation can be done on the digits to confirm that the number is valid. The magnetic strip on the back of the driver's license usually includes the document number.

Document numbers were not available from the data broker websites mentioned earlier; however, several offers on the darknet include physical copies of actual driver's licenses and data compiled from swipes of the magnetic strips of driver's licenses. These range from $1 to $10 per license for copies and bulk discounts for magnetic strip data of $0.01 to $1 each. Using this information, an attacker could learn whether these numbers are encoded values derived from other information or sequentially assigned, and could then predict document numbers for other driver's licenses without purchasing more data. Or, an attacker could use the purchased data to opportunistically impersonate Texas and New York voters. In this writing, we assume that all references to driver's license data for New York and Texas includes the document number at the cost of $0.01 per record unless stated otherwise or is obvious from context.

In short, darknet markets sell all the kinds of personal information an attacker would need to impersonate voters on the 36 websites. For example, for about $1,000, an attacker could have acquired the names, addresses, dates of birth, gender, and SSNs of most adult Americans in the United States by purchasing two datasets. Driver's license numbers with issue dates, names, addresses, and dates of birth were available for as little as $0.41 per record, with negotiable volume discounts or bundled in some offers with SSN data, though not necessarily available for every driver in America. An attacker who wanted to use voter lists to target voters to impersonate could have purchased statewide voter lists for Alaska, Nevada, and Rhode Island for $64 on the darknet instead of $1,020 from named sources.

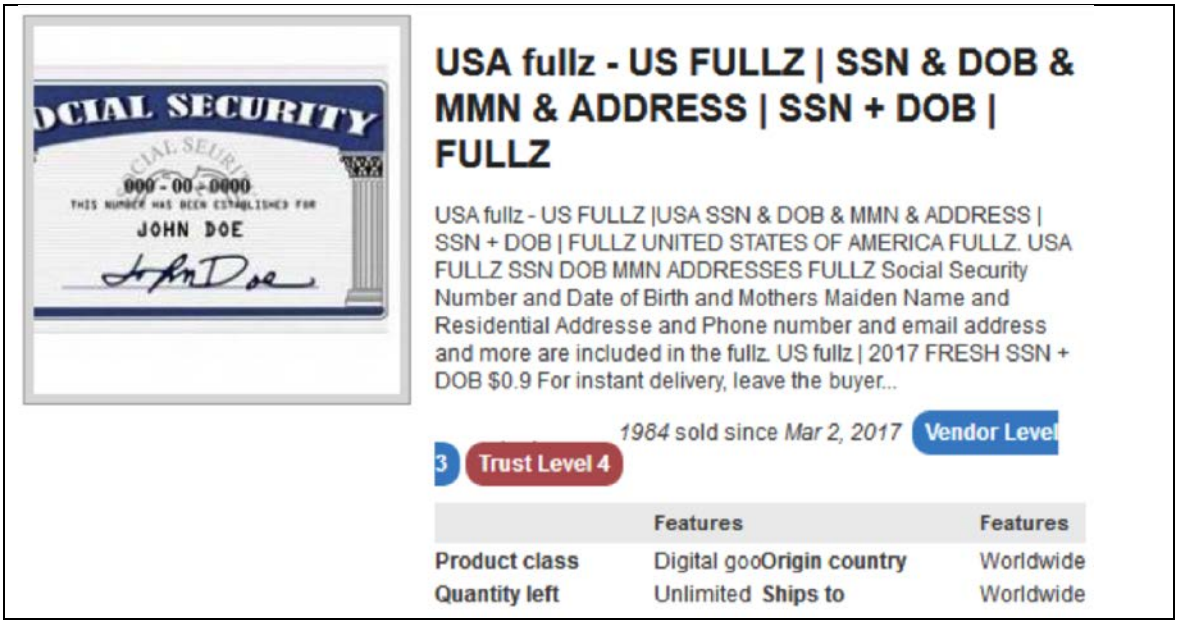


Figure 11. Excerpt of a 2017 darknet offer of a dataset containing name, address, date of birth, and SSN for $0.90 per record. Volume discounts available.



Figure 12. Excerpt of a darknet offer of a state dataset containing name, address, date of birth, and SSN. Similar solicitations were found for each of the 50 states and the District of Columbia.

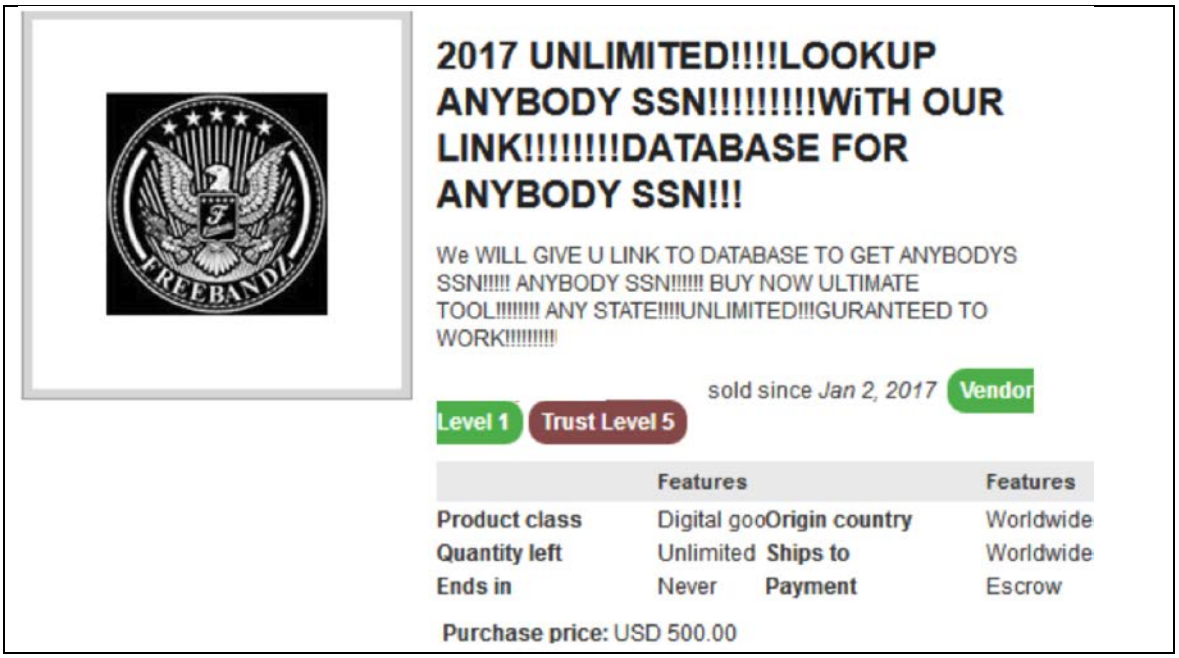
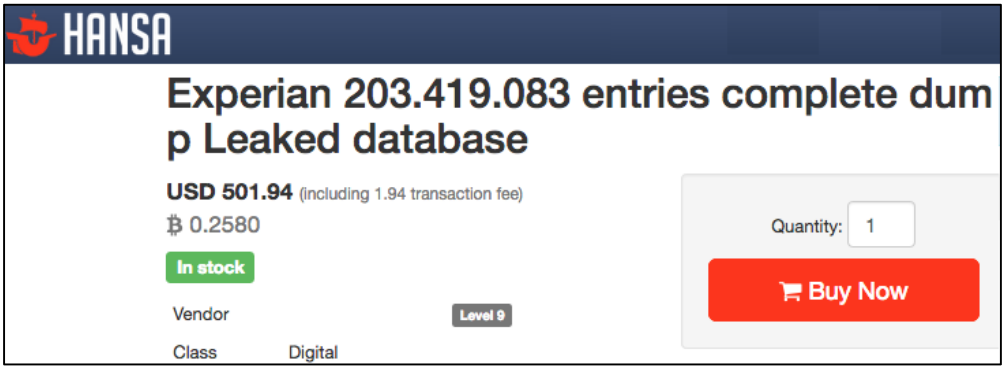


Figure 13. Excerpt of a darknet offer of unlimited access to a dataset containing name, address, date of birth, and SSN for $500. Ability to download the entire dataset included. The offer claims that 100 percent of SSNs appear.



(a)

FIRST NAME
MI
LAST NAME
PREFIX
ADDRESS
SUITE/APT
CITY
STATE
ZIP5
ZIP4
DELIVERY POINT BAR CODE
FIPS STATE CODE
FIPS COUNTY CODE
LATITUDE
LONGITUDE
ADDRESS TYPE INDICATOR
0 = Undetermined
1 = Single Family Dwelling
2 = Apartment with unit designator
3 = Apartment without unit designator
4 = Rural Route
5 = Post Office Box
COMMUNITY REINVESTMENT ACT (CRA)
1 = LOW INCOME
2 = MODERATE INCOME
3 = MIDDLE INCOME
4 = HIGH INCOME
TELEPHONE NUMBER
ETHNIC CODE
ETHNIC-CONFIDENCE
ETHNIC-GROUP
LANGUAGE
RELIGION
HISPANIC COUNTRY-CODE
ASSIMILATION CODES
CREDIT RATING
A = 800+
B = 750 - 799
C = 700 - 749
D = 650 - 699
E = 600 - 649
F = 550 - 599
G = 500-549
H = under 499

DO NOT CALL FLAG
PROPERTY TYPE
A = Single Family Dwelling Unit
B = Condo
C = Cooperative
D = 2-4 Unit (Duplex, Triplex, Quad)
E = Miscellaneous Residence (Combo Store/Fla
G = Apartment (5+ Units)
M = Mobile Home
T = Timeshare

GENDER CODE
U = Unknown
F = Female
M = Male
INFERRED HOUSEHOLD RANK (Values 1 - 9
HOME OWNER SOURCE FLAG
V = Verified Home Owner
H = Highly Likely Home Owner
P = Probably Home Owner
DATE OF BIRTH - YEAR
DATE OF BIRTH - MONTH
DATE OF BIRTH - DAY
EXACT AGE - BASED OFF DOB
INFERRED AGE
INCOME - ESTIMATED HOUSEHOLD
A = Under $10,000
B = $10,000 - $14,999
C = $15,000 - $19,999
D = $20,000 - $24,999
E = $25,000 - $29,999
F = $30,000 - $34,999
G = $35,000 - $39,999
H = $40,000 - $44,999
I = $45,000 - $49,999
J = $50,000 - $54,999
K = $55,000 - $59,999
L = $60,000 - $64,999
M = $65,000 - $74,999
N = $75,000 - $99,999
O = $100,000 - $149,999
P = $150,000 - $174,999
Q = $175,000 - $199,999
R = $200,000 - $249,999
S = $250,000 +
NET WORTH
A = Less than $1
B = $1 - $4,999
C = $5,000 - $9,999
D = $10,000 - $24,999
E = $25,000 - $49,999
F = $50,000 - $99,999
G = $100,000 - $249,999
H = $250,000 - $499,999
I = Greater than $499,999
CREDIT - RANGE OF NEW CREDIT
0 = $0 - $100
1 = $101 - $300
2 = $301 - $500
3 = $501 - $1,000
4 = $1,001 - $3,000
5 = $3,001 - $5,000
6 = $5,001 - $9,999
7 = Greater than $9,999
PRESENCE OF CHILDREN
Y= Yes
MARITAL STATUS IN THE HHLD
M = Married
S = Single
A = Inferred Married
B = Inferred Single
HOME OWNER / RENTER
O = Home Owner
R = Renter

(b)

Figure 14. Excerpts of (a) darknet offer for a dataset containing (b) fields that included the names, addresses, gender, and dates of birth of 203 million Americans. Cost for a copy of the dataset was $501.94.

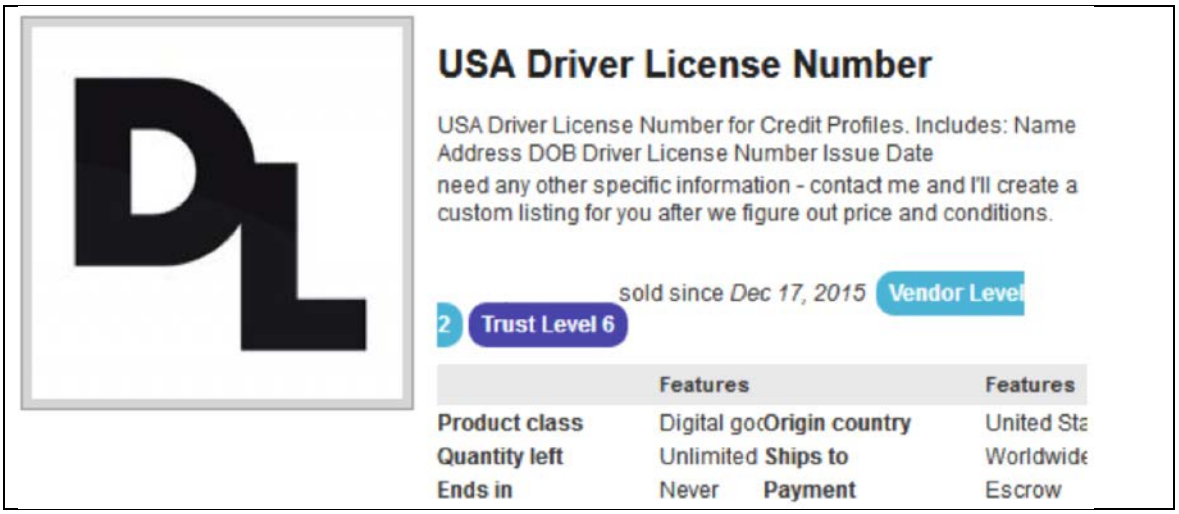


Figure 15. Excerpt of a darknet offer of access to a dataset containing name, address, date of birth, driver's license number, and driver's license issuance date. Bulk prices seem negotiable but the total number of driver licenses available is not clear.



Figure 16. Excerpt of a darknet offer for access to a dataset containing the name, address, date of birth, SSN, and driver's license number of people in the United States.

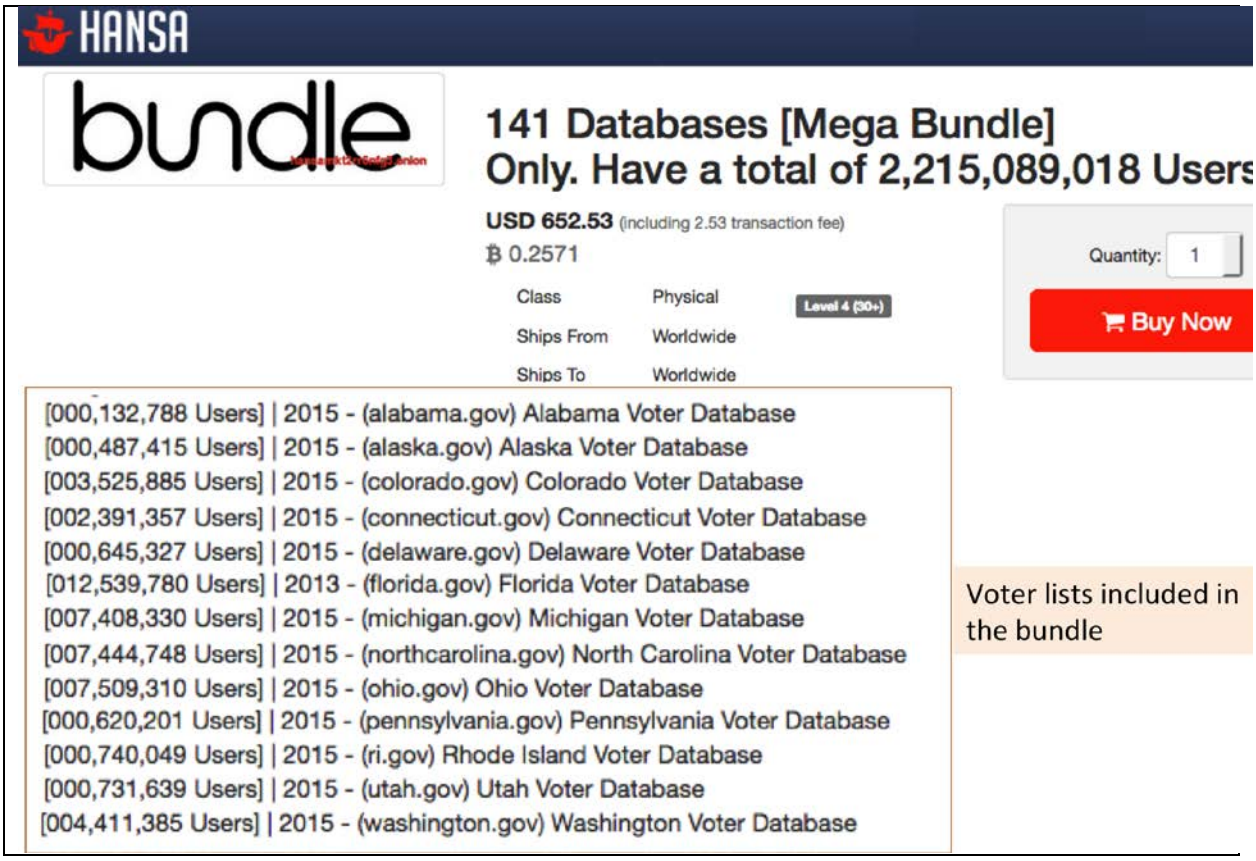
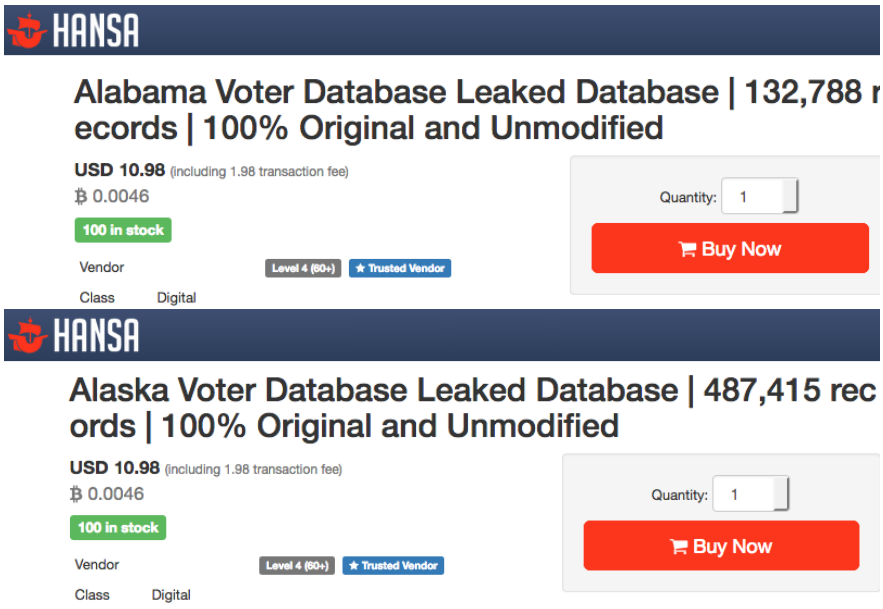
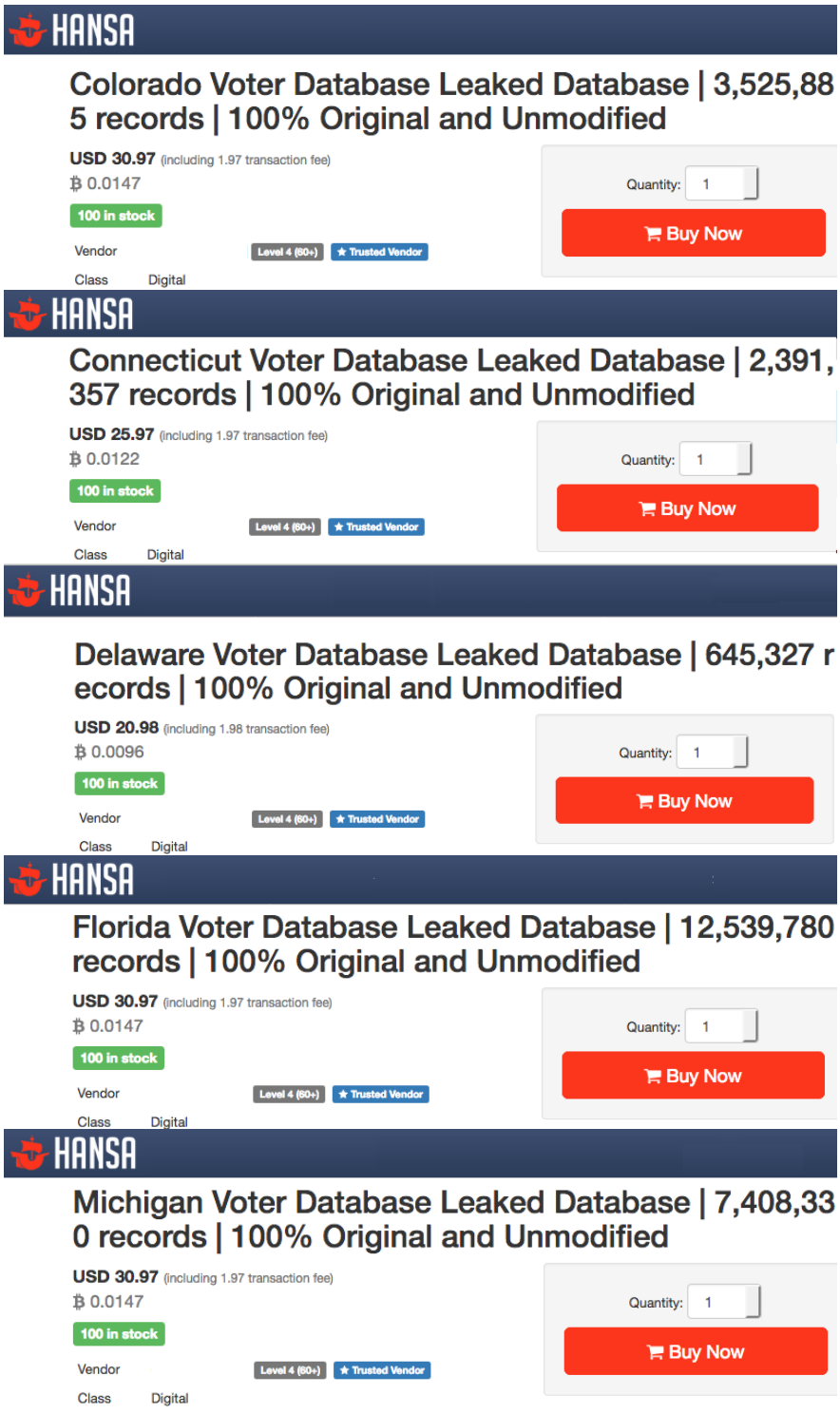


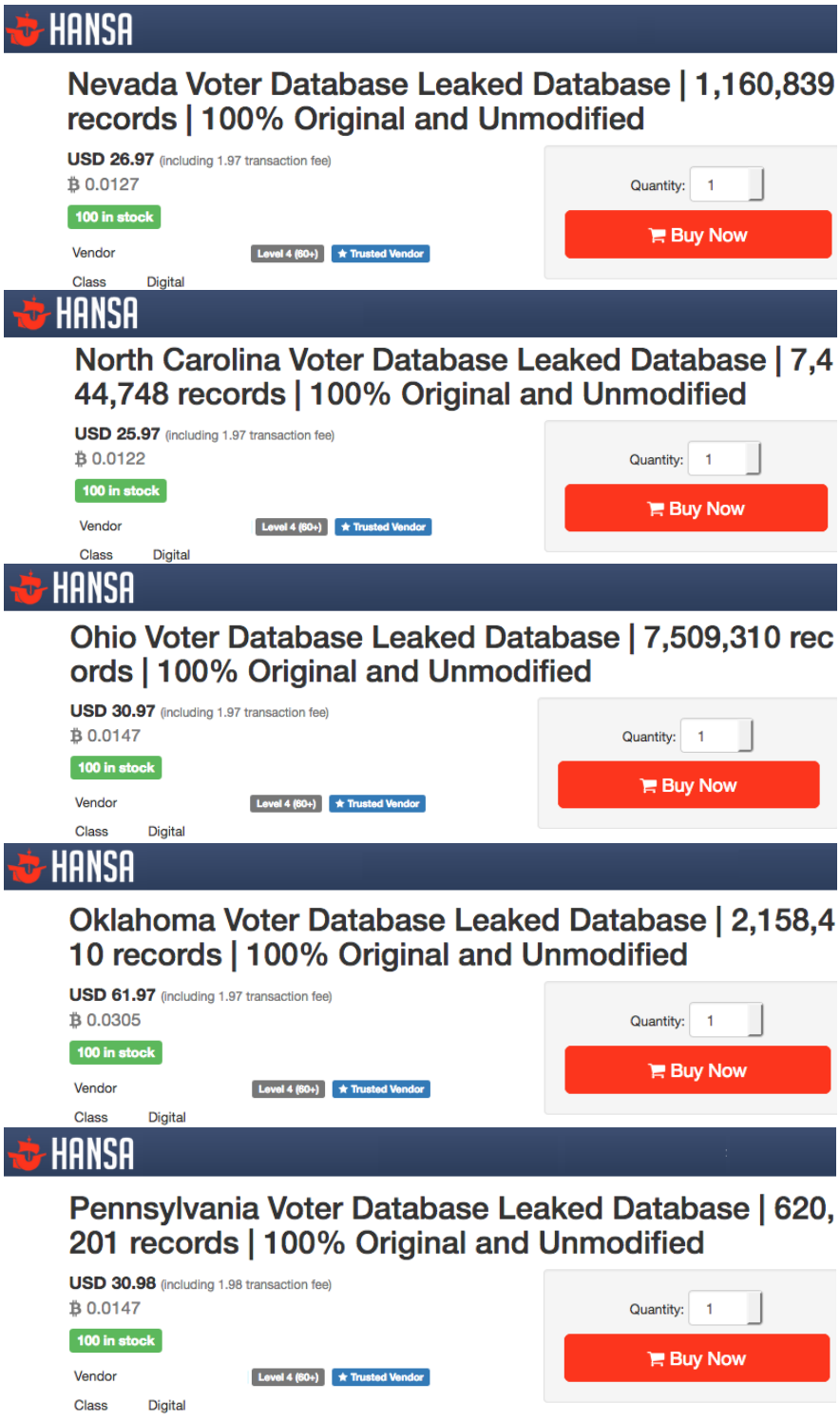
Figure 17. Excerpt of a darknet offer for a dataset containing 141 databases for $653, including 2015 voter files from 12 states. Some appear to include 100 percent of the voters and others a fraction of the voters in that state. Florida's list is for 2013.



Sweeney L, Yoo J, Zang J. Voter Identity Theft: Submitting Changes to Voter Registrations Online to Disrupt Elections. *Technology Science*. 2017090601. Version 2. September 6, 2017. http://techscience.org/a/2017090601



Sweeney L, Yoo J, Zang J. Voter Identity Theft: Submitting Changes to Voter Registrations Online to Disrupt Elections. *Technology Science*. 2017090601. Version 2. September 6, 2017. http://techscience.org/a/2017090601



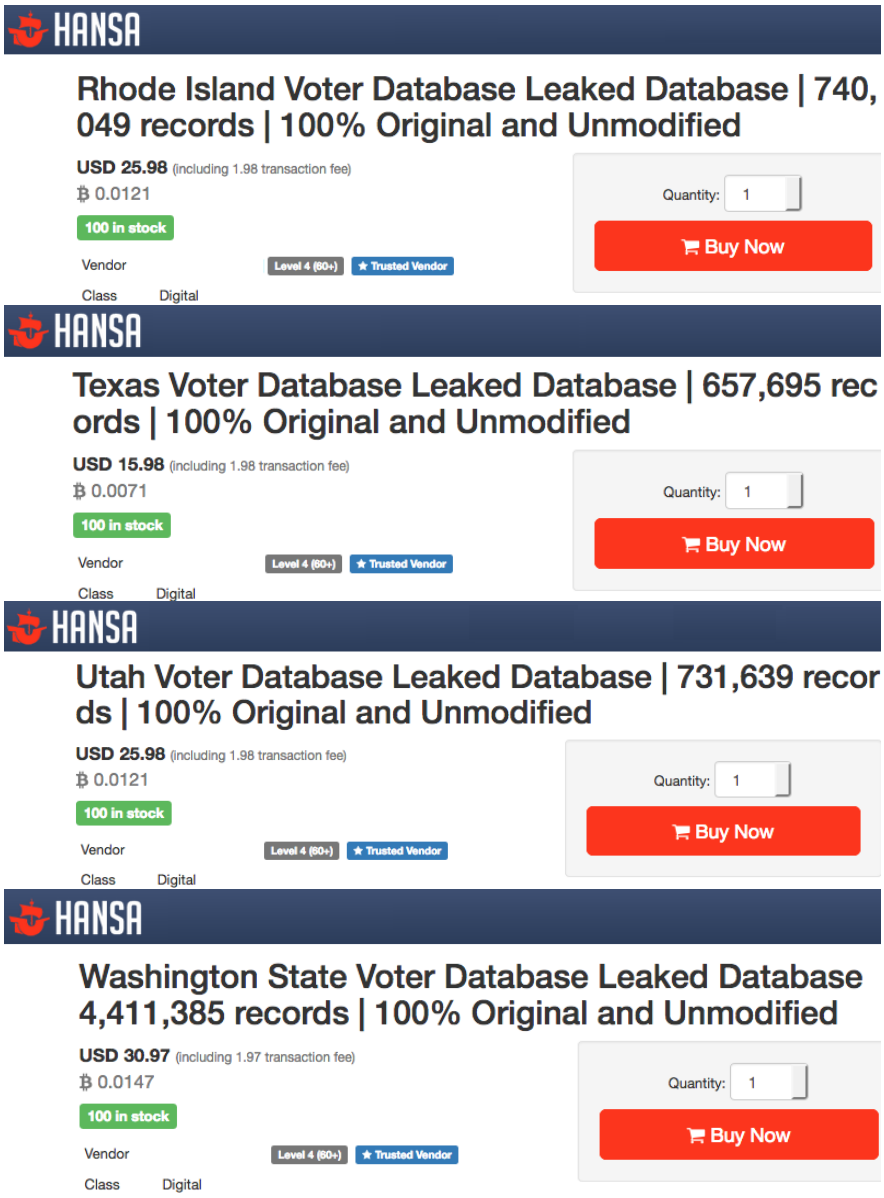
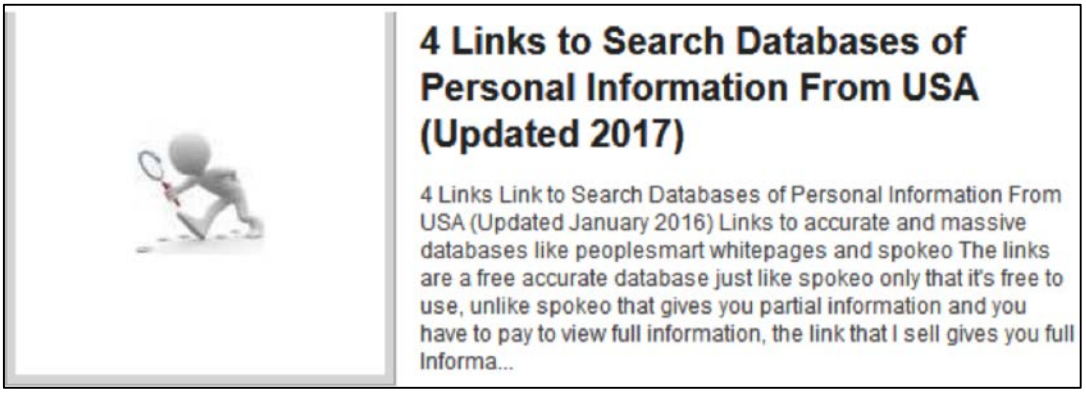


Figure 18. Excerpts of darknet offers for voter lists by state. These offers included lists for 13 of the 36 states in our study, of which 9 included all voters (Alaska, Colorado, Connecticut, Delaware, Michigan, Nevada, Ohio, Rhode Island, and Washington) and 4 did not include all the voters in the state (Alabama, Pennsylvania, Texas, and Utah). Offers shown for voter lists from states not in our study are from Florida, North Carolina, and Oklahoma.



Figure 19. Excerpt of a darknet offer for discounted access to popular public data broker websites.



(a)



(b)

Figure 20. Excerpt of 2 darknet offers for URL links to personal data claiming (a) a source similar to a popular public data broker and (b) the SSNs of 70 percent of all Americans having SSNs.

Results for Step 3: Data Sources: (g) Deep Web

The public often learns about large data breaches of sensitive personal information when computer security investigators find datasets publicly available at openly accessible website URLs (e.g. [65],[51]). No password or account credentials block access. Instead, the datasets

are available to anyone who happens to know the URL, discovers the URL, or perhaps purchases knowledge of the URL from a darknet market. These breached data contain the names, addresses, dates of birth, SSNs, and driver's license numbers of millions of Americans (e.g., [33], [34], [35], [36]).

There are three primary reasons these data could be unguarded. First, a legitimate data holder may naïvely believe that the URL is so well hidden that no one beyond those with whom they share the URL will know the dataset is there. Meanwhile, the open access allows them to easily share voluminous information with business associates and others.

Second, the data holder may use weak security settings on the cloud storage platform that hosts the data. These settings may even default to open access and then remain unchanged because the data holder is unaware of the vulnerability.

Lastly, an insider or a thief could share the dataset at an obscure URL link on purpose, enabling outsider access without actually having personal possession of the data. As mentioned earlier, the open access makes it easier to share large amounts of information.

The last of these scenarios seems the most likely for an attacker seeking to impersonate voters at the 36 websites.

Earlier we described the breaches of millions of voter records from Illinois and Arizona government election offices [41], [42], [43], [44]. In this subsection, we turn our attention to political campaigns.

We surveyed breach notices, news articles, and the websites of data analytic companies that work with political campaigns to identify the kinds of personal information that could be leaked to, discovered or stolen by an attacker.

Overall, we found that political campaigns hold a treasure trove of personal data on Americans. They have up-to-the-minute voter lists, often embellished with lifestyle and interest information. A campaign often holds credit card information of those who make donations. Data analytic companies working with or on behalf of campaigns may use driver's license data to identify unregistered voters and Social Security numbers to disambiguate and link information across disparate datasets. Resulting voter profiles may be shared back to the campaigns in aggregate or detailed form. In large data operations, a campaign may maintain all the demographic and government-issued data an attacker needed to impersonate voters.

We also found reports of ongoing attempts to steal and acquire personal information on Americans from political campaigns, and some political campaign datasets were found openly available on the deep web.

NBC News reported that U.S. authorities traced massive political cyber-espionage activities targeting the 2008 U.S. presidential campaigns of Barack Obama and John McCain to the

People's Republic of China [143]. The goal, according to the officials, was to export massive amounts of internal data from both campaigns.

An article published in *Time* magazine described ongoing hacking attempts faced constantly by the U.S. presidential campaigns of Barack Obama and Mitt Romney in 2012 [144]. According to the article, organized crime attempted to steal credit card data, and there were ongoing attempts to gain access to all kinds of valuable data seemingly by foreign nation states. Zac Moffatt, digital director of the Romney campaign, reported being "under constant attack" with highly visible incidents occurring "four or five times a week." Other organizations under attack included the National Republican Congressional Committee.

Data from political campaigns have also been found publicly and openly available on the deep web. For example, a national voter list containing 191 million U.S. voter registration records appeared openly available on the Web in December 2015 [65]. The 300 GB dataset contained names, addresses, dates of birth, party affiliations, and voting history for each voter in all 50 states and the District of Columbia.

Another example is 170 GB of 2016 political profiles on American voters found freely available on the Web [51]. The dataset contained virtually every voter in the United States, including all the kinds of information found on voter lists embellished with surveys and detailed analyses about enthusiasm for Trump.

In short, the voter lists needed to impersonate voters at the 36 websites are openly and freely available on the deep web, and data breaches containing many of the needed government identifiers have been reported.

Results for Step 4. Record and Test CAPTCHAs.

Of the 36 websites in our study, 11 (31 percent) had some form of CAPTCHA. These states were Connecticut, Delaware, Indiana, Nebraska, Nevada, Ohio, Oregon, Pennsylvania, South Carolina, Utah, and Washington.

Connecticut's website had the easiest CAPTCHA to defeat. All of the 100 images we captured consisted of a 5-digit number displayed using the same font, print angle, and background. Figure 21 shows some examples. Here are three ways that an attacker could semi- or fully automate responses to these CAPTCHAs. First, in a preliminary step, the attacker could copy a separate image of each digit, 0 to 9, and then write a simple image processing program to match the stored images of digits against a CAPTCHA image of a 5-digit number to report which digits appear. This approach allows the attacker to fully automate change of address changes at scale.

As an alternative, the attacker could construct a dictionary that associates each 5-digit number, 0000 to 99999, with its corresponding image. Each time the website displays a new 5-digit code, the attacker enters the code by hand, recording the image and the answer. Any

repeated CAPTCHA images would then be recognized automatically, so this semi-automated approach has slow manual performance in the beginning until enough recognized images re-appear.

As a third and final alternative, the attacker could write a simple computer program that sends a copy of the CAPTCHA image to a network of human helpers, (e.g., Amazon Mechanical Turk [58]), who in turn provide the text answer back to the program, which then enters the text as a response to the CAPTCHA. This semi-automated approach would have been possible because the Connecticut website's page did not reset even after 5 minutes, which is more than enough time to route the image to human helpers and get an answer back. Amazon's Mechanical Turk costs about $0.042 per CAPTCHA (about $0.03 to the human worker plus Amazon's 40 percent overhead) [145], [146].

Ohio's website had a similar but more advanced CAPTCHA. All of the 100 images we captured showed 5 characters. Each image had a different font effect and print angle against a random dot background. Figure 3a and Figure 22 show examples. Researchers describe ways to automate the recognition of this kind of CAPTCHA, and they report a 30 to 50 percent success rate [59]. Of course, the semi-automated means described earlier using a network of human helpers would defeat this CAPTCHA because the pages did not reset even after 5 minutes.

Delaware, Indiana, South Carolina, Utah, and Washington's websites used Google's reCAPTCHA v1 system of images of street signs from Google Street view [147]. Figure 3b and Figure 23 show examples. As mentioned earlier, the Google Street View algorithm [61] and a program from a commercial company [60] are automated ways to defeat these CAPTCHAs. These web pages also accepted answers after 5 minutes of idle time, so the semi-automated approach using human helpers would have defeated them also.

Lastly, Nebraska, Nevada, Oregon, and Pennsylvania's websites use Google's reCAPTCHA v2 which uses an "I'm not a robot button" based on user click behavior rather than an image-based quiz [148]. After the "I am not a Robot" box is clicked (Figure 24), a 3x3 grid of images appears, to which the user clicks an answer by selecting those images having a specific characteristic (Figure 3c and Figure 3d). These CAPTCHAs require a response of mouse clicks within the grid of images in only 2 minutes, so a network of human helpers would likely not work.

However, both the ReCAPTCHA v1 and ReCAPTCHA v2 services allowed an audio alternative to visual images. The attacker could have used semi-automated or automated approaches to defeat "verbal CAPTCHAs" even when image CAPTCHAs could not be defeated. Clicking the headphones icon replaces the image CAPTCHA with a verbal one (Figure 24). A verbal CAPTCHA plays a sound file of a person speaking some digits or words. The user responds by typing the spoken words into the text box (Figure 24). The attacker could automate this sound-to-text test by writing a computer program that sends the sound file to a speech recognition program (e.g., Google's own speech API [149]), which in turn, returns the resulting

text to enter as a response to the CAPTCHA (see the release of a computer program after the election that worked in this manner [150], which forced Google to make further changes to the ReCAPTCHA service). Alternatively, the attacker could develop a semi-automated solution in which a network of human helpers sends text responses to verbal prompts. The wait time for responses was 2 minutes, so performance of the human network would have to be optimized.

In short, 11 (31 percent) of the 36 websites had a CAPTCHA service that attempted to limit the rate at which an attacker could change addresses on the website, but an attacker could have defeated all of them. Connecticut, Delaware, Indiana, Ohio, South Carolina, and Washington's websites had CAPTCHAs that fully automated computer programs could defeat. Ohio's website was the easiest to semi-automate, because the image could be farmed out in real-time to an online network of human helpers who sent back the text to enter into the CAPTCHA. Using a network of human helpers would have cost about $0.042 per CAPTCHA. Nebraska, Nevada, Oregon, and Pennsylvania's websites used the same, very difficult to defeat image CAPTCHA service. However, the CAPTCHAs on these websites, as well as those on the websites of Delaware, Indiana, South Carolina, and Washington, included a verbal option, which could be defeated by automated or semi-automated means by converting sound to text using a speech recognition program or sending the sound file to a network of human helpers who sent back the text from the spoken words.

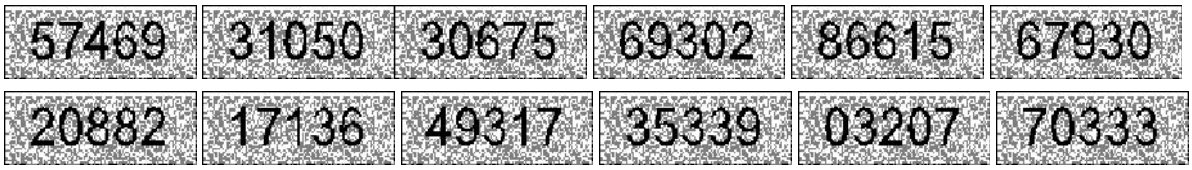


Figure 21. Twelve sample CAPTCHA images from the Connecticut website.

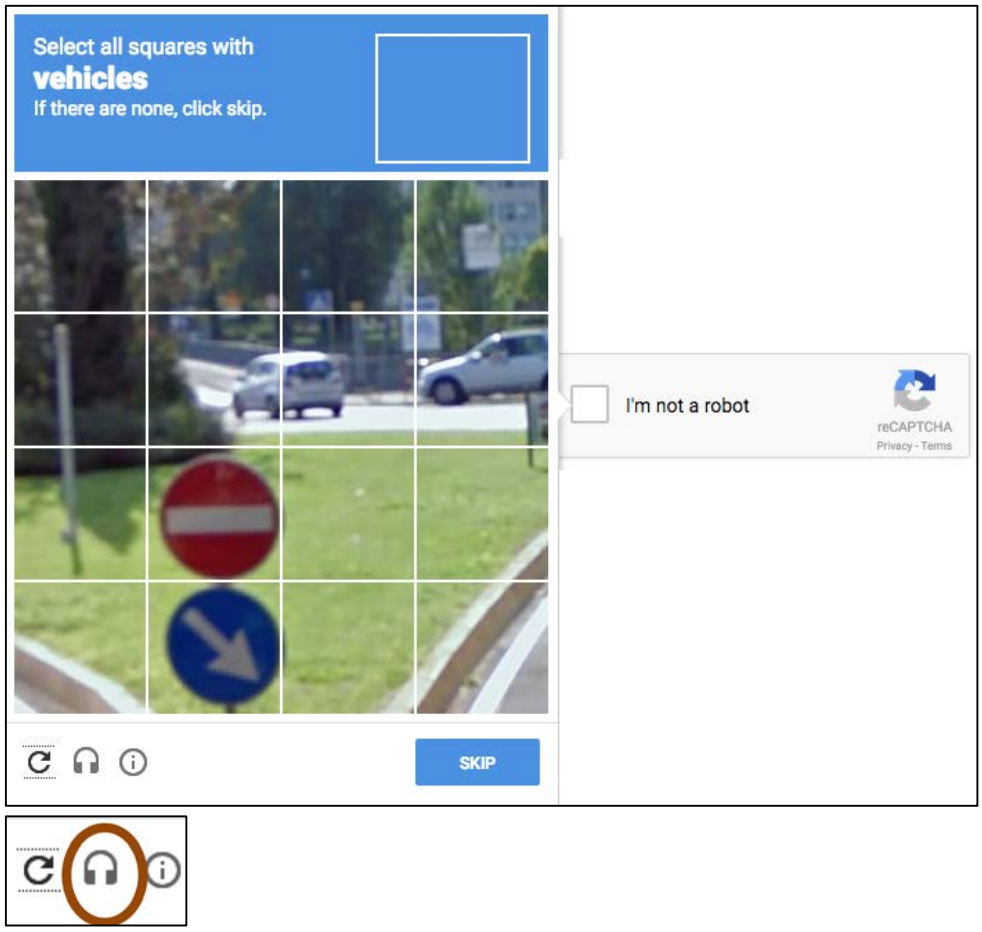


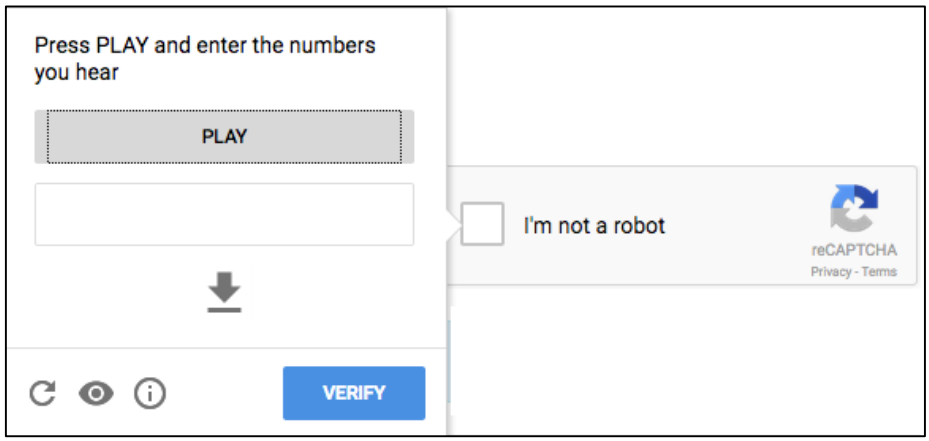
Figure 22. Eight sample CAPTCHA images from the Ohio website.





Figure 23. Sample CAPTCHA images from the websites of Delaware (top row), Indiana, South Carolina, Utah, and Washington (bottom row).





(c)

Figure 24. Sample reCAPTCHA image (a), with verbal option (c) selected by clicking on the headphones icon (b).

Results for Step 5. Test Automation.

In 2016, we wrote computer programs using the Python programming language that automatically submitted names and demographic information to 19 voter websites. We randomly chose the websites for Alabama, Alaska, Connecticut, District of Columbia, Delaware, Hawaii, Indiana, Kansas, Kentucky, Massachusetts, Minnesota, Nevada, New York, Rhode Island, Utah, Vermont, Washington, and West Virginia. These websites reported a voter's polling place and were not the same websites used for registering or changing the registrations of voters. We made this distinction to further insulate our automation tests from actual voter rolls. Like the websites used for changing addresses, polling place websites required entering information into fields on a web page and "clicking" a form submission button. Often the submission process spanned multiple web pages.

We used the Python language and constructed different kinds of programs to best match different Python libraries to the features present on websites. Here are several simple examples.

Figure 25a shows the web form that finds polling places in Ohio from street addresses. Figure 25b shows a simple Python program for looking up one voter. By writing a program that automatically changed the values associated with 'frmstnum' and 'frmstname' to be all the addresses found in the voter file for Ohio, we learned all the polling places associated with voter addresses.

Similarly, Figure 26a shows an example of the website that finds polling places in Alabama. Figure 26b shows a simple Python program for looking up one voter. By changing values for 'nameLast,' 'dobMonth,' 'dobDay,' and 'dobYear,' our program found all polling places for voter addresses in Alabama.

The other websites required more advanced programming, but we were completely successful at automating searches on all 8 state websites using Python programs we wrote.

[We merged some of our results with results from other approaches to create a public service website voteGPS.org, where voters could easily find their polling place in those states whose information was not as readily available in early 2016. By November 2016, however, polling place information was readily available online.]

In short, we demonstrated how simple Python programs can be used to automate online activity at state websites.



(a)

```
#Load the form variables with values and show the result.
import mechanize
from mechanize import Browser

url="http://www.voterfind.com/lickingoh/pollfinder.aspx"
br = mechanize.Browser()
response = br.open(url)
br.select_form(nr=0)
br.form['frmstnum']='999'
br.form['frmstname']='Middleground'
br.submit()
print br.response().read()
```

(b)



(c)

Figure 25. Ohio website features: (a) form for looking up polling places in an Ohio county [http://www.voterfind.com/lickingoh/pollfinder.aspx] and (b) a simple, complete Python program to look up the polling place for one Ohio voter at the county website. The code was further modified to look up the polling places of all voters in the county and then modified further (c) for all voters in the state [http://voterlookup.sos.state.oh.us/voterlookup.aspx].



(a)

```
import urllib2
import urllib

url="https://myinfo.alabamavotes.gov/VoterView/PollingPlaceSearch.do"

values = dict(countyRequired="true", action="search", county="300100",
    nameLast="      ", dobMonth="02", dobDay="04", dobYear="1961",
    fullElectionListLong="22515_200000,03/01/2016 2016 PRES PREF PRIMARY

try:
    req = urllib2.Request(url, urllib.urlencode(values))
    response_page = urllib2.urlopen(req).read()
except urllib2.HTTPError, details:
    print "error found."
print response_page
```

(b)

Figure 26. Alabama website (a) for looking up polling places [https://myinfo.alabaavotes.gov/VoterView/PollingPlaceSearch.do] and (b) a simple, complete Python program to look up the polling place of one Alabama voter at this website. The code was further modified to look up the polling places of all voters in the state.

Results for Step 6. Compare Total Attack Costs.

We have now shown that the primary ingredients an attacker needed to execute a change of address attack at scale on any of the 36 state websites in 2016 were: (1) computing facilities

to edit voter records online; (2) personal data to impersonate voters; (3) computing facilities to defeat CAPTCHAs; and, (4) the programs or programming skills to automate tasks. We have also shown variability among the state websites in terms of data requirements and CAPTCHAs, so in this section we compare costs of executing a change of address attack by website.

We focus on perpetrating attacks at scale using automation. The computer programs needed to impersonate voters or to defeat CAPTCHAs differ by website, so a scaled attack requires programming. For the remainder of this section, we assume the attacker had the programming skills needed to write the kinds of straightforward Python or equivalent programs required (e.g., Figure 25b and Figure 26b) or had a confederate with that skill. Our cost comparisons do not include programming costs.

If an attacker sought to cause widespread havoc by changing all or most voter records at a state website, the impact would have been noticed at the polling places. But what if the attacker wanted to have impact but remain undetected? In this case, he would seek to alter records strategically across the state in small percentages so that no one polling place would have a large number of disenfranchised voters complaining. The overall impact of this approach across the polling places in the state could be considerable. A change of 1, 2, 5, or perhaps even 10 percent of the voter rolls may not be noticed. So we look at the costs of a change-of-address attack on the 36 websites using different sources for personal data and different ways of addressing CAPTCHAs to disenfranchise 1, 5 and 10 percent of the voters in the state.

Results for Step 6a. Total Attack Costs Using Named Sources

In this subsection, we compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using personal data from named sources using an automated method to defeat CAPTCHAs.

Changes to 1 percent of the voters in a state would have affected anywhere from 505 voters in Alaska, which had the fewest voters, to 170,283 voters in California, the largest population of voters. The median was 33,000 voters and the average 42,550, with a standard deviation of 39,951. Changes to 5 and 10 percent of the voters equated to 2,526 and 5,052 Alaskan voters and 851,415 and 1,702,829 Californian voters, respectively. See Figure 28.

A computer can work around the clock with possibly some down time for programming or storage changes. In our computations, we estimate a computer working 20 hours a day for 30 days a month, for a monthly total of 600 hours.

Extrapolating from our experiment with automated retrieval of polling place data (prior subsection), we assume that a computer could impersonate a voter on a state website and make an address change within 1 minute. In other words, changing the addresses of 505 Alaskan voters would take a computer 505 minutes; done using a single computer, it would

take less than a day. Figure 27 shows the number of machine hours needed to edit the addresses of 1 percent of the voters on each state website. The minimum is 8 hours to change the 505 voters on Alaska's website. The maximum is 2,838 hours to change the 170,283 voters on California's website. The median is 33,000 hours and the average 42,550 hours.

An attacker can use a farm of computers as needed by purchasing virtual cloud machines (e.g., [151]). We use the current price of a basic configuration from Microsoft Azure, which is $13 per machine per month. We compute the machine cost per state based on the machine time needed to edit the voter rolls. This is the number of required machine hours divided by 600 hours/month times $13/month for machine rental.

We start by ignoring any machine costs needed to respond to CAPTCHAs or harvest personal information from broker websites. Instead, we first look at the computer time needed to enter personal information at the website. Then, in a few paragraphs below, we add the additional costs.

For example, to change 1 percent of the Alaskan voters, the machine cost for editing the website is 8 hours divided by 600 times $13, or $0.18. For California, the cost is 2838 hours divided by 600 times $13, or $61. Figure 27 enumerates the machine costs for editing voter rolls by state to change the records of 1 percent of the voters. The median cost is $12 and the average $15, with a standard deviation of 14. The total machine costs to edit the records of 1 percent of voters at all 36 websites would have been $553, the cost of about 1 month of computing by 43 cloud machines.

An attacker could use virtual machines to scale attacks as speed and resources permitted. For example, the 2,838 hours needed to edit voter registrations on the California website could be done by a single computer working almost 5 months, 5 virtual machines working almost 1 month, or 10 virtual machines working for about 2 weeks, all for the same $61 in computing costs.

As described in detail earlier, changing voter addresses requires access to voter names and demographics (Table 3). Prices and sources vary. Suppose the attacker purchases voter data from government sources for all purchases less than $100 and from brokers or government sources for all other states (Table 8). In that case, voter data for 16 states cost less than $100 each and for 6 states, the cost is the maximum of $2,000 each; the total for all 36 states is $17,679. The median cost is $114 and the average $491, with a standard deviation of $733. Figure 27 lists the voter costs for per state.

An attacker would also need government-issued identifiers such as SSNs and driver's license data. As reported earlier, we found named data brokers that provided SSNs and driver's license data; the least expensive was a flat $40 per month with an additional $0.01 for detailed data that included the driver's license issue date.

Texas and New York websites need document numbers. We budget $0.01 for each document number based on an assortment of possibilities. An attacker could purchase driver's license data on the darknet that includes document numbers $0.01 each. An attacker may have derived a scheme for predicting document numbers in these states, as discussed earlier, in which case, there are no additional data costs but there are additional computational costs for multiple tries. So, for convenience, we use one cent cost per record to model the costs of these different possibilities.

In estimating costs, we applied the $40 per state for each 600 hours of computer time. This cost model accommodates seeming limits on the "unrestricted" access to be what a single computer can access within a month from the broker's website –i.e., 600 hits per month. So, the cost to acquire government data differs by state website based on the number of hits per month. We used the number of machine hours computed earlier for each state, divided by 600, and multiplied by 40. This computation appears for each state in Figure 27 as the cost of government data (Gov ID). There is no cost shown for Delaware because no government number was needed on its website. Otherwise, the least expensive was $1 for Alaska. The most expensive was $1,678 for Texas because of the one-cent lookup cost for the driver's license document number. The median was $40 and the average $170, with a standard deviation of $381. The cost of acquiring all the SSNs and driver's license data needed to impersonate 1 percent of the voters on the 36 websites totaled $5,958. Figure 27 lists per state totals.

Using a data broker website to display the SSNs and driver's license data would require a computer program to look up the information and harvest the needed numbers. We used the same 1 minute-per-voter time estimate, so the computing costs are the same as those for editing the voter information online. These values are replicated in Figure 27 as the machine cost for acquiring personal data.

Some websites used CAPTCHAs, but these could be automated. For these websites, we computed the machine costs for running the program to respond to the CAPTCHA to also be 1 minute per voter, with the exception of Ohio, where we used 3 minutes per voter because of the algorithm's hit rate being 30 to 50 percent. These values are stored as the machine costs for CAPTCHAs on Figure 27. The minimum was $2 for Delaware. The maximum was $85 for Ohio. The median was $9 and the average $18, with a standard deviation of $24. The total machine cost for defeating all CAPTCHAs automatically was $195.

In short, an attacker could acquire personal data from government sources and data brokers and use programs to defeat CAPTCHAs to change the addresses of 1 percent of the voters on the 36 websites. Figure 27 summarizes the costs of doing so. Delaware was the least expensive at $5, and Texas was the most expensive at $3,059. The median cost for an attacker to attack one state is $339; the average is $692 with a standard deviation of $841. The voter addresses of 1 percent of the voters on all 36 websites could be changed for a total of $24,926.

The cost of acquiring voter data from government offices and data brokers dominated the total cost to change 1 percent of the voters on all 36 websites. Specifically, the cost of the voter data was $17,679 (71 percent) of the $24,926 total. The acquisition of SSNs and driver's license data totaled $5,958 (23 percent) and was the second largest cost.

As would be expected, costs increase with the number of voter records changed per state. As we just itemized, changing 1 percent of the voter records on all 36 websites would have cost $24,926. Changing 5 percent of the voter records on all 36 websites would roughly double the cost to $53,915, and changing 10 percent would increase the cost to $90,152. See Figure 28.

State	1% of Voters	Website Editing (Machine)		Personal Data from Named Sources			CAPTCHA		Relative Cost	
		Hours	Cost	Voter	Gov ID	Machine	Machine	Human		
AL	33,308	555	$12	$2,000	$37	$12			$2,061	8.3%
AK	505	8	$0	$20	$1	$0			$21	0.1%
AZ	35,885	598	$13	$2,000	$40	$13			$2,066	8.3%
CA	170,283	2,838	$61	$30	$189	$61			$342	1.4%
CO	35,711	595	$13	$0	$40	$13			$65	0.3%
CT	21,154	353	$8	$0	$24	$8	$8		$46	0.2%
DE	6,820	114	$2	$0			$2		$5	0.0%
GA	46,957	783	$17	$250	$52	$17			$336	1.3%
HI	7,069	118	$3	$500	$8	$3			$513	2.1%
IL	80,451	1,341	$29	$500	$894	$29			$1,452	5.8%
IN	48,292	805	$17	$2,000	$54	$17	$17		$2,106	8.4%
KS	16,446	274	$6	$200	$18	$6			$230	0.9%
KY	31,053	518	$11	$450	$35	$11			$507	2.0%
LA	30,318	505	$11	$2,000	$34	$11			$2,056	8.2%
MD	39,001	650	$14	$128	$43	$14			$200	0.8%
MA	42,718	712	$15	$0	$47	$15			$78	0.3%
MI	74,546	1,242	$27	$0	$83	$27			$137	0.5%
MN	32,693	545	$12	$46	$36	$12			$106	0.4%
MO	41,909	698	$15	$100	$47	$15			$177	0.7%
NE	11,407	190	$4	$500	$13	$4	$4		$525	2.1%
NV	16,864	281	$6	$500	$19	$6	$6		$537	2.2%
NJ	58,193	970	$21	$3	$65	$21			$109	0.4%
NM	11,980	200	$4	$2,000	$13	$4			$2,022	8.1%
NY	124,933	2,082	$45	$0	$1,388	$45			$1,478	5.9%
OH	78,610	1,310	$28	$0	$87	$28	$85		$229	0.9%
OR	25,686	428	$9	$500	$29	$1	$9		$548	2.2%
PA	87,223	1,454	$31	$20	$97	$31	$31		$211	0.8%
RI	7,709	128	$3	$0	$9	$3			$14	0.1%
SC	27,223	454	$10	$160	$302	$10	$10		$492	2.0%

State	1% of Voters	Website Editing (Machine)		Personal Data from Named Sources			CAPTCHA		Relative Cost	
		Hours	Cost	Voter	Gov ID	Machine	Machine	Human		
TX	151,011	2,517	$55	$1,272	$1,678	$55			$3,059	12.3%
UT	15,584	260	$6	$0	$17	$6	$6		$34	0.1%
VT	4,698	78	$2	$0	$5	$2			$9	0.0%
VA	55,297	922	$20	$2,000	$61	$20			$2,101	8.4%
WA	42,703	712	$15	$0	$474	$15	$15		$521	2.1%
DC	4,798	80	$2	$0	$5	$2			$9	0.0%
WV	12,749	212	$5	$500	$14	$5			$523	2.1%
Totals	**1,531,787**	**25,530**	**$553**	**$17,679**	**$5,958**	**$542**	**$195**	**$0**	**$24,926**	
			2%	**71%**	**24%**	**2%**	**1%**	**0%**		

Figure 27. Total attack costs for changing the addresses of 1 percent of the voters on the 36 websites using automated programs to respond to CAPTCHAs and purchasing voter lists from government sources when those lists cost less than $100 or else, for all other states, from the least expensive named source (Table 8). Estimates include 1 minute per voter for a computer to make the address change on the state website ("Editing (Machine)"), another minute to scrape the SSN and driver's license number from the data broker's website, and another minute to respond to the CAPTCHA (except 3 minutes for Ohio because the program had only 30-50 percent accuracy). Machine time cost was estimated at $13 for 600 hours (a computer work month). SSN and driver's license numbers purchased from a data broker for a flat fee of $40 per 600 hours plus $0.01 for each driver's license issue date. In the cases of NY and TX, we presume that additional driver's license data were inferable.

State	1% of Voters			5% of Voters			10% of Voters		
	#Voters	Costs	% Costs	#Voters	Costs	% Costs	#Voters	Costs	% Costs
AL	33,308	$2,061	8.3%	166,540	$2,305	9.1%	333,080	$2,611	10.5%
AK	505	$21	0.1%	2,526	$25	0.1%	5,052	$29	0.1%
AZ	35,885	$2,066	8.3%	179,423	$2,329	9.2%	358,847	$2,658	10.7%
CA	170,283	$342	1.4%	851,415	$1,591	6.3%	1,702,829	$3,152	12.6%
CO	35,711	$65	0.3%	178,556	$327	1.3%	357,113	$655	2.6%
CT	21,154	$46	0.2%	105,772	$232	0.9%	211,543	$464	1.9%
DE	6,820	$5	0.0%	34,098	$25	0.1%	68,195	$49	0.2%
GA	46,957	$336	1.3%	234,785	$680	2.7%	469,570	$1,111	4.5%
HI	7,069	$513	2.1%	35,345	$565	2.2%	70,689	$630	2.5%
IL	80,451	$1,452	5.8%	402,257	$5,260	20.7%	804,514	$10,020	40.2%
IN	48,292	$2,106	8.4%	241,462	$2,530	10.0%	482,924	$3,060	12.3%
KS	16,446	$230	0.9%	82,231	$351	1.4%	164,461	$502	2.0%
KY	31,053	$507	2.0%	155,267	$735	2.9%	310,535	$1,019	4.1%

State	1% of Voters			5% of Voters			10% of Voters		
	#Voters	Costs	% Costs	#Voters	Costs	% Costs	#Voters	Costs	% Costs
LA	30,318	$2,056	8.2%	151,590	$2,278	9.0%	303,181	$2,556	10.3%
MD	39,001	$200	0.8%	195,005	$486	1.9%	390,009	$843	3.4%
MA	42,718	$78	0.3%	213,592	$392	1.5%	427,184	$783	3.1%
MI	74,546	$137	0.5%	372,728	$683	2.7%	745,455	$1,367	5.5%
MN	32,693	$106	0.4%	163,463	$346	1.4%	326,926	$645	2.6%
MO	41,909	$177	0.7%	209,547	$484	1.9%	419,094	$868	3.5%
NE	11,407	$525	2.1%	57,037	$625	2.5%	114,074	$750	3.0%
NV	16,864	$537	2.2%	84,320	$685	2.7%	168,641	$870	3.5%
NJ	58,193	$109	0.4%	290,964	$536	2.1%	581,928	$1,069	4.3%
NM	11,980	$2,022	8.1%	59,898	$2,110	8.3%	119,795	$2,220	8.9%
NY	124,933	$1,478	5.9%	624,663	$7,392	29.1%	1,249,325	$14,784	59.3%
OH	78,610	$229	0.9%	393,051	$1,146	4.5%	786,103	$2,293	9.2%
OR	25,686	$548	2.2%	128,430	$739	2.9%	256,860	$977	3.9%
PA	87,223	$211	0.8%	436,114	$977	3.8%	872,228	$1,934	7.8%
RI	7,709	$14	0.1%	38,544	$71	0.3%	77,088	$141	0.6%
SC	27,223	$492	2.0%	136,114	$1,820	7.2%	272,229	$3,480	14.0%
TX	151,011	$3,059	12.3%	755,054	$10,207	40.1%	1,510,109	$19,142	76.8%
UT	15,584	$34	0.1%	77,921	$171	2.6%	155,842	$342	1.4%
VT	4,698	$9	0.0%	23,492	$43	0.2%	46,983	$86	0.3%
VA	55,297	$2,101	8.4%	276,487	$2,507	9.9%	552,974	$3,014	12.1%
WA	42,703	$521	2.1%	213,514	$2,604	10.2%	427,027	$5,207	20.9%
DC	4,798	$9	0.0%	23,990	$44	0.2%	47,980	$88	0.4%
WV	12,749	$523	2.1%	63,744	$617	2.4%	127,489	$734	2.9%
Totals	**1,531,787**	**$24,926**	**100%**	**7,658,937**	**$53,915**	**100%**	**15,317,874**	**$90,152**	**100%**

Figure 28. Total attack costs for changing the addresses of 1, 5, and 10 percent of the voters on the 36 websites using automated programs to respond to CAPTCHAs and purchasing voter lists from government sources when those lists cost less than $100 or else, for all other states, from the least expensive named source (Table 8). Estimates include 1 minute per voter for a computer to make the address change on the state website, another minute to scrape the SSN and driver's license number from the data broker's website, and another minute to respond to the CAPTCHA (except 3 minutes for Ohio because the program had only 30-50 percent accuracy). Machine time cost was estimated at $13 for 600 hours (a computer work month). SSN and driver's license numbers purchased from a data broker for a flat fee of $40 per 600 hours plus $0.01 for each driver's license issue date. In the cases of NY and TX, we presume that additional driver's license data were inferable.

Results for Step 6b. Total Attack Costs using Darknet Sources

In this sub-section, we compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using personal data from darknet sources with automated means to defeat CAPTCHAs.

An attacker could use personal datasets sold on darknet markets in strategic ways to impersonate voters at scale. One way would be to combine the SSN dataset that contained the names and SSNs of most Americans (Figure 13) with the dataset that contained the name, address, and demographics of 203 million Americans (Figure 14) to make a combined "National Dataset." The cost for the two datasets together would have been $1,002, and the contents would complete the requirements for 6 states and the ELSE requirements for 5 states. See path A in Figure 29.

The information from the National Dataset could then be used with the state website for Nebraska to learn needed driver's license numbers for Nebraska voters at no extra cost; see path C on Figure 29. The combined information could be used with computer programs to compute needed driver's license number from demographics for Maryland, Michigan, and New Jersey, at no extra cost; see path D in Figure 29.

The website for Texas required a voter ID number (Table 2). The voter ID numbers were present on Texas voter lists purchased from named sources, as well as from darknet sources. The darknet offer was for a partial list of Texas voters and cost $15.98. The statewide dataset from Texas itself was $1,272. We used the statewide version in our analyses because of its completeness. We also use the $0.01 per record model cost for the document or "audit" number as described earlier.

Finally, an attacker could use additional information from the data broker described earlier to learn needed driver's license numbers and issuance dates for 26 states at a cost of $40 per machine month and $0.01 per issuance date; see path B in Figure 29.

An attacker may have made a deal with a supplier of driver license data for a dataset that included the driver's license numbers and issuance dates of most Americans. We found an offer on a darknet market that seemed to encourage the negotiation of such a dataset (Figure 15), but we did not contact the supplier or engage in such negotiations so we do not use the acquisition of the dataset in these computations. If an attacker did acquire driver's license data in bulk from a darknet source, the total attack cost could be much lower than what we report. See path X on Figure 29.

Figure 30 itemizes the relative costs to change 1 percent of the voters on each of the 36 state websites using the National Dataset and driver's license data from a data broker, the Texas voter list from a government office in Texas, a state website, and algorithms that infer driver's license numbers; see paths A, B, C, and D in Figure 29. There are no state-specific costs for

voter data other than Texas because all voter (and non-voter) information is included in the $1,002 cost for the National Dataset.

The National Dataset is not voter-specific, so we estimate the number of attempts (or "tries") required at the state website to find registered voters. The U.S. Census reports the percentage of each state's residents who are registered voters [152]. The number of tries per change is 1 divided by the percentage of voters in the state. Nationally, the proportion of registered voters is 64.2 percent, which increases the number of attempts by a factor of 1.6. In other words, for every three people's information attempted, two would likely be voters. The "Tries" column in Figure 30 reports how this plays out for each state. The factor is 1.0 for Texas because we use the actual Texas voter list. The smallest factor, other than for Texas, is 1.3 for Washington, D.C., where 75.9 percent of the population are registered voters. The largest factor is 2 for Hawaii, which has less than half of its population registered to vote. The average and median is 1.5 with a standard deviation of 0.2.

The number of attempts is the number of voter records sought to be changed times the number of tries. For example, a 1 percent change in Alaska would involve changing 505 voter records. With 69.1 percent of Alaska's population registered to vote, the factor is 1.4. We estimate that it will take about 707 attempts to edit 505 voter registrations, because 505 times 1.4 is 707.

These additional attempts to locate voters through state websites increase the number of hours machines would have to spend editing voter rolls. We still assume that each attempt takes 1 minute. The number of machine hours to change 1 percent of the voter records in a state is the number of required machine hours divided by 600 hours/month times $13 per month for machine rental. See the machine cost column under "Website Editing" in Figure 30 for state-level totals. Alaska has the lowest cost, $0.26, for 12 hours of effort. California has the highest at $114. The median is $16 and the average $24, with a standard deviation of $27. The total for all 36 states is $587.

The only state-specific cost for personal information was the Texas voter list and New York and Texas document numbers. There were no personal data costs for any other states or for SSNs, only the $1,002 cost of the National Dataset itself. Driver's license data, however, is not included in the National Dataset. The least expensive data broker for driver's license data charged a flat $40 per month with an additional $0.01 for detailed data that included the driver's license issue date. As we did in the previous assessment, we applied the $40 per state for each 600 hours of computer time. The total cost for acquiring driver's license data was $6,109. The average cost among those states needing driver's license data to be purchased was $255 with a standard deviation of $460. The median was $52. See the "Driver ID" column on Figure 30 for costs per state.

We also estimated 1 minute for each access to the driver's license data at the broker's website. This increased the machine costs for those states requiring driver's license data that

could not be inferred from demographics or SSNs or looked up for free. The total was $587. The median was $16 and the average $24, with a standard deviation of $27. See the machine costs for "Personal Data" in Figure 30 for totals by state.

As in the prior assessment, we estimate the costs of automated approaches to defeat any website CAPTCHAs. We allocate 1 minute per attempt for each state having a CAPTCHA except for Ohio, where we estimate 3 minutes per attempt. The per state costs appear under the "CAPTCHA" heading on Figure 30. The minimum is $4 for Delaware. The maximum is $123 for Ohio. The median is $14 and the average $26, with a standard deviation of $34. The total machine costs for defeating all CAPTCHAs automatically were $284.

Overall, an attacker could use the National Dataset to change 1 percent of the voter records at the 36 websites for a total cost of $10,081. Changing 5 percent of the voter records on all 36 websites cost $41,310, and changing 10 percent cost $80,347.

An attacker could have changed voter records using the National Dataset, but for any state other than Texas, targeting specific voters could only be done based on residence, race inferred from first name for Blacks and last name for Hispanics and Asians, or gender. If an attacker wanted to impersonate voters based on actual rather than inferred party affiliation, then voter lists would be necessary for the other states.

An alternative to using the National Dataset is to join the SSN dataset with the least expensive options for acquiring statewide voter lists from named and darknet sources. The SSN dataset cost $500 (Figure 13), and the cost of acquiring statewide voter lists using the least expensive sources totaled $$17,196. Combining these datasets would complete the requirements for 6 states and the ELSE requirements for 5 states. See Path E in Figure 29. We computed the per state costs for using the combination of statewide voter lists and the SSN dataset. The total cost to change 1 percent of the voter records on all 36 websites was $20,073. Changing 5 percent of the voter records cost $29,605, and changing 10 percent cost $41,488.

In short, two datasets that contained the names, demographics and SSNs of most adult Americans were available from darknet sources for $1,002 total. Together, these datasets provide a national dataset that dramatically lowered the costs for an attacker by avoiding the purchase of voter lists altogether. For $10,081, an attacker could use the two datasets and a data broker to change 1 percent of the voter records on the 36 websites. Had the attacker acquired a dataset that contained the SSNs of most Americans and statewide voter lists from darknet sources, the cost to change 1 percent of the voter records would have been $20,073. Both options cost less than the $24,926 option described earlier, which used data primarily from named government and data broker sources.

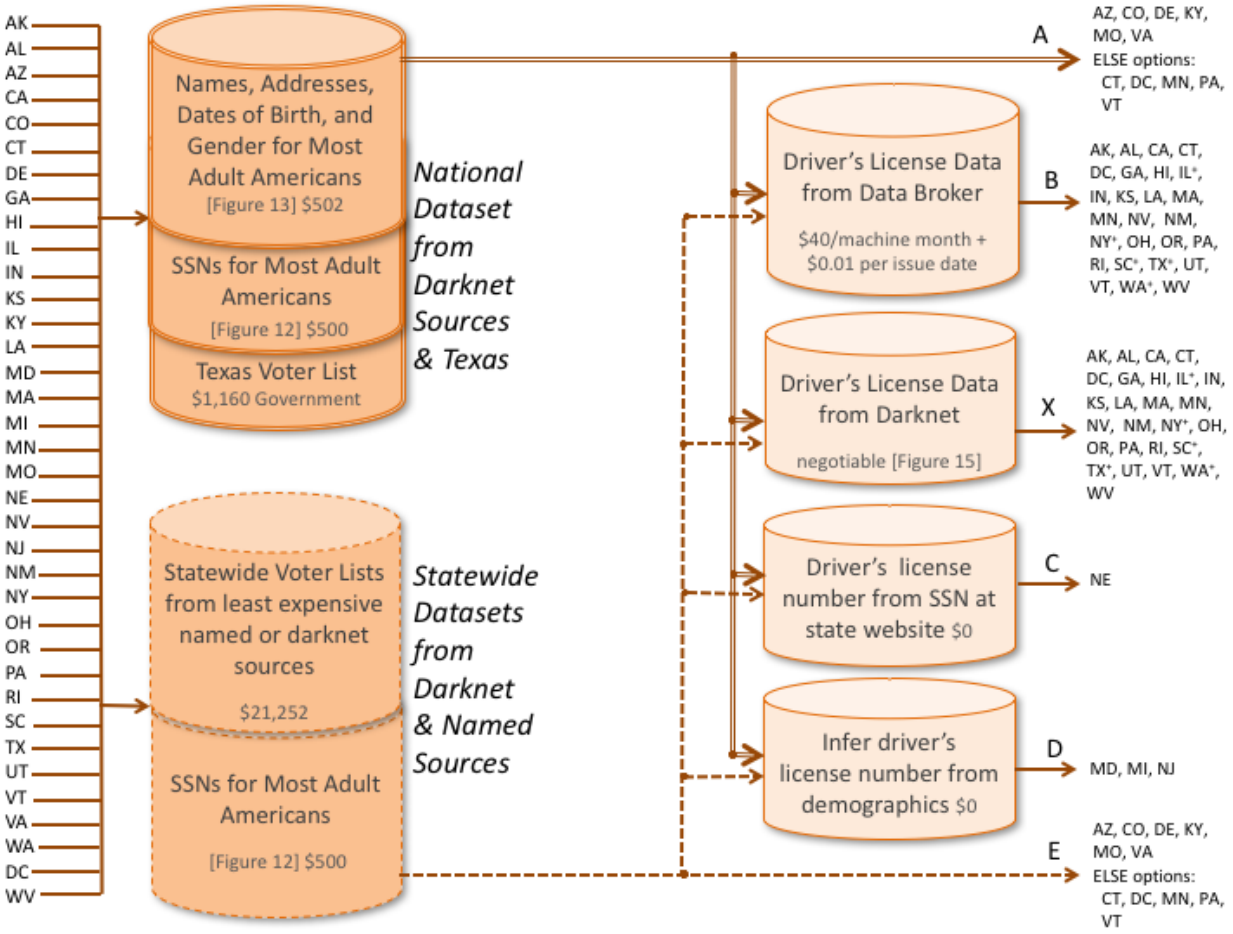


Figure 29. Six paths for an attacker to have acquired the personal data needed to impersonate voters on each of the 36 websites: (a) combination of two datasets from the darknet that jointly produce a National Dataset having the names, demographics and SSNs of most adult Americans combined with the Texas Voter List purchased from Texas; (e) an alternative of statewide voter lists and SSNs combined; and then, addition of driver's license data from (b) a data broker, (c) a state website, (d) inferred from demographics, or, (x) a dataset having driver's license numbers and issuance dates acquired from the darknet (e.g., Figure 15). The (+) superscript identifies those states whose websites required driver's license issue date information.

State	1% of Voters	Website Editing (Machine Cost)			Personal Data			CAPTCHA	Relative Cost	
		Tries	Hours	Cost	Profiles, SSN	Driver ID	Machine	Machine		
AL	33,308	1.5	816	$18	National	$54	$18		$90	0.9%
AK	505	1.4	12	$0	National	$1	$0		$1	0.0%
AZ	35,885	1.7	989	$21	National		$21		$21	0.2%
CA	170,283	1.9	5275	$114	National	$352	$114		$580	5.8%

State	1% of Voters	Website Editing (Machine Cost)			Personal Data			CAPTCHA	Relative Cost	
		Tries	Hours	Cost	Profiles, SSN	Driver ID	Machine	Machine		
CO	35,711	1.5	873	$19	National		$19		$19	0.2%
CT	21,154	1.6	552	$12	National	$37	$12	$12	$73	0.7%
DE	6,820	1.5	170	$4	National			$4	$7	0.1%
GA	46,957	1.6	1221	$26	National	$81	$26		$134	1.3%
HI	7,069	2.0	237	$5	National	$16	$5		$26	0.3%
IL	80,451	1.5	1957	$42	National	$935	$42		$1,020	10%
IN	48,292	1.5	1218	$26	National	$81	$26	$26	$160	1.6%
KS	16,446	1.5	408	$9	National	$27	$9		$45	0.4%
KY	31,053	1.5	769	$17	National		$17		$17	0.2%
LA	30,318	1.4	716	$16	National	$48	$16		$79	0.8%
MD	39,001	1.5	966	$21	National		$21		$21	0.2%
MA	42,718	1.5	1045	$23	National	$70	$23		$115	1.1%
MI	74,546	1.4	1743	$38	National		$38		$38	0.4%
MN	32,693	1.4	747	$16	National	$50	$16		$82	0.8%
MO	41,909	1.4	969	$21	National				$21	0.2%
NE	11,407	1.4	265	$6	National			$6	$11	0.1%
NV	16,864	1.6	458	$10	National			$10	$20	0.2%
NJ	58,193	1.6	1598	$35	National				$35	0.3%
NM	11,980	1.7	337	$7	National	$22	$7		$37	0.4%
NY	124,933	1.7	3529	$76	National	$1,485	$76		$1,638	16%
OH	78,610	1.4	1885	$41	National	$126	$41	$123	$330	3.3%
OR	25,686	1.5	635	$14	National	$42	$1	$14	$71	0.7%
PA	87,223	1.4	2101	$46	National	$140	$46	$46	$277	2.7%
RI	7,709	1.6	200	$4	National	$13	$4		$22	0.2%
SC	27,223	1.4	658	$14	National	$316	$14	$14	$359	3.6%
TX	151,011	1.0	2517	$55	$1,272	$1,678	$55		$3,059	30%
UT	15,584	1.5	389	$8	National	$26	$8	$8	$51	0.5%
VT	4,698	1.4	112	$2	National	$7	$2		$12	0.1%
VA	55,297	1.4	1328	$29	National				$29	0.3%
WA	42,703	1.4	1018	$22	National	$495	$22	$22	$561	5.6%
DC	4,798	1.3	105	$2	National	$7	$2		$12	0.1%
WV	12,749	1.6	334	$7	National				$7	0.1%
National Dataset					$1,002				$1002	
Totals	**1,531,787**		**38,152**	**$827**	**$2,274**	**$6,109**	**$587**	**$284**	**$10,081**	
				8%	**23%**	**61%**	**6%**	**3%**		

Figure 30. Total attack costs for changing the addresses of 1 percent of the voters on the 36 websites using automated programs to respond to CAPTCHAs and acquiring from darknet sources a national dataset of SSNs (Figure 13) and a national dataset of profiles having the

names, address, gender, and dates of birth of 203 million Americans (Figure 14). Estimates include 1 minute per attempt for a computer to make the address change on the state website ("Website Editing"), where the number of attempts is based on a factor derived from the percentage of registered voters in the state population ("Tries"). Costs include a minute to scrape driver's license number from the data broker's website, and another minute to respond to the CAPTCHA (except 3 minutes for Ohio because the program had only 30-50 percent accuracy). Machine time estimated at $13 for 600 hours (a computer work month). Driver license numbers purchased from a data broker for a flat fee of $40 per 600 hours plus $0.01 for each driver's license issue date. In the cases of NY and TX, we presume that additional driver's license data were inferable. Costs for driver's license data could be reduced if attacker acquired a dataset of driver's license data from the darknet (e.g., Figure 15).

Results for Step 6c. Total Attack Costs Using Deep Web Sources

In this sub-section, we compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using personal data from deep web sources with automated means to defeat CAPTCHAs.

Had an attacker accessed complete datasets either found on the Internet or shared secretly from a confederate, then there would have been no costs for data at all. However, machine costs to edit the voter records at the website and to defeat CAPTCHAs would remain. These costs total $748 to change 1 percent of the voter records at all 36 websites. Costs would jump to $3,739 and $7,477 to change 5 and 10 percent of the records, respectively.

Results for Step 6d. Total Attack Costs Using Semi-Automated Means to Defeat CAPTCHAs

In this sub-section, we compute the total attack costs to change 1 percent, 5 percent, and 10 percent of voter records using a network of human helpers to defeat CAPTCHAs.

In our prior assessments, we used computer programs to defeat CAPTCHAs because these automations were applicable to the websites as they appeared in 2016. Months after the election, however, Google modified its reCAPTCHA v2 service to make it harder to use speech recognition for sound-to-text automation. That means that an attacker in 2016 had an advantage that an attacker today does not. Instead of automated response, an attacker today might need to turn to a network of human helpers.

With a network of human helpers, costs increase on a per-CAPTCHA basis. We use a rate of $0.042 per CAPTCHA to pay humans for simple image-to-text CAPTCHAs and $0.084 per CAPTCHA for more difficult sound-to-text CAPTCHAs. We contextualize these increased costs in the two scenarios we described earlier: (a) changing 1 percent of the voter records using personal data from named sources (Figure 27), and (b) changing 1 percent of the voter

records using a national datasets of personal data from darknet sources (Figure 30). Actual costs vary by the kind of CAPTCHAs found at the 11 state websites that had CAPTCHAs.

The kind of CAPTCHA that appears on Connecticut's website (Figure 21) is not part of Google's service, so it remains susceptible to the same automatic matching of digit images that we discussed earlier. We estimate its cost, as we did previously, based on the allocated machine time for the program to respond, namely, 1 minute per CAPTCHA. In the assessment that used data from named sources, processing CAPTCHAs for 353 Connecticut voters cost $8. This cost rises to $12 in the assessment that used the National Dataset because each change requires 1.6 attempts.

Ohio's website had a kind of CAPTCHA (Figure 22) that was also not part of Google's service. The computer program used to solve this kind of CAPTCHA provided a correct answer 30 to 50 percent of the time, meaning that often it would make multiple attempts per voter. A network of human helpers could help. Instead of multiple tries by the program, the program could try first, and if not successful, forward the replacement CAPTCHA to the network of human helpers. There would be at most 2 CAPTCHA attempts per voter. We set machine costs based on 6 minutes per voter to handle the extra processing. We expect half of the attempts to be forwarded to the human helpers, so the human costs are $0.042 times half the number of voters. In the assessment that used data from named sources, the machine costs to handle CAPTCHAs for 78,610 voters is $170 and the costs for human helpers is $1,651. These costs rise to $245 in machine costs and $1,651 in human helper costs in the assessment that used the National Dataset because it required 1.6 machine attempts per change.

Delaware, Indiana, South Carolina, Utah, and Washington's websites used Google's reCAPTCHA v1 system of images of street signs (Figure 23). Automation was done with a version of the Google street view algorithm. The automation appears to still work at the time of this writing, but we explore a semi-automated alternative. The attacker's computer forwards the images to human helpers, who return the text read from the images. We base machine costs on an estimate of 5 minutes per CAPTCHA to allow for the handling, and human costs on $0.042 per CAPTCHA. Figure 31 shows the totals for the two scenarios for each of these states.

Finally, Nebraska, Nevada, Oregon, and Pennsylvania's websites used Google's reCAPTCHA v2, which uses clicks on images but offers a verbal option. Automation can be done by using a speech recognition program to respond to the verbal CAPTCHA option. In March, following the November election, Google made the speech more difficult to discern, which increased the number of errors from speech recognition. A network of human helpers could help. Human helpers would receive a sound file and then return the text heard from the sound. We based machine costs on an allocation of 2 minutes per CAPTCHA and human helper costs on $0.084 per CAPTCHA. We used the higher rate for humans because Google has made the task more difficult for humans as well as for speech recognition software. Figure 31 shows the totals for the two scenarios for each of these states.

With semi-automated approaches, total costs increase as the number of voters increases because the costs paid to the human helpers are per CAPTCHA; see Figure 31. Using personal data from darknet sources for the same state website always had higher CAPTCHA costs than using personal data from named sources because this approach involved more attempts to authenticate registered voters; see Figure 31.

Costs of addressing CAPTCHA become a greater percentage of the total attack cost as the number of voter records increases. Consider the assessment that used automated CAPTCHA responses and personal data from named sources. The total automated CAPTCHA costs to change 1 percent of voter records across all 36 websites were $195 (0.8 percent of $24,926 total), costs to change 5 percent of voters were $973 (1.8 percent of $53,915), and costs to change 10 percent were $1,946 (2.2 percent of $90,152). Now consider the assessment that used semi-automated CAPTCHA responses and the same data. Changing 1 percent of voter records across all 36 websites cost $19,416 (43 percent of $44,681 total), changing 5 percent of voters cost $97,080 (64 percent of $152,692), and changing 10 percent cost $194,161 (67 percent of $287,705).

Moving from automation to semi-automation of CAPTCHAs increased the total attack cost for the 36 websites by $19,755 (from $24,926 to $44,681) when changing 1 percent of the voter records using data from named sources and by $29,006 (from $10,081 to $39,087) when using data from darknet sources. See Figure 32.

In short, a website that uses CAPTCHAs can dramatically increase the cost of an attack if the attacker has no means other than labor to circumvent the CAPTCHA. In 2016, however, automation was possible on all the state websites that had CAPTCHAs, rendering CAPTCHAs less of a deterrent and keeping attack costs low.

State/ Percentage of Voters/Number of Voters			Automated CAPTCHA Costs		Semi-Automated CAPTCHA Costs			
			Named	Darknet	Named Sources		Darknet Sources	
			Machine	Machine	Machine	Human	Machine	Human
CT	**1%**	21,154	$8	$12	$8	$0	$12	$0
	5%	105,772	$38	$60	$38	$0	$60	$0
	10%	211,543	$76	$120	$76	$0	$120	$0
DE	**1%**	6,820	$2	$4	$12	$286	$18	$429
	5%	34,098	$12	$18	$62	$1,432	$92	$2,144
	10%	68,195	$25	$37	$123	$2,864	$184	$4,288
IN	**1%**	48,292	$17	$26	$87	$2,028	$132	$3,069
	5%	241,462	$87	$132	$436	$10,141	$660	$15,343
	10%	482,924	$174	$264	$872	$20.283	$1,319	$30,685
NE	**1%**	11,407	$4	$6	$8	$958	$11	$1,336
	5%	57,037	$21	$29	$41	$4,791	$57	$6,682
	10%	114,074	$41	$57	$82	$9,582	$115	$13,364

State/ Percentage of Voters/Number of Voters			Automated CAPTCHA Costs		Semi-Automated CAPTCHA Costs			
			Named Machine	Darknet Machine	Named Sources		Darknet Sources	
					Machine	Human	Machine	Human
NV	1%	16,864	$6	$10	$12	$1,417	$20	$2,307
	5%	84,320	$30	$50	$61	$7,083	$99	$11,536
	10%	168,641	$61	$99	$122	$14,166	$198	$23,071
OH	1%	78,610	$85	$123	$170	$1,651	$245	$2,375
	5%	393,051	$426	$613	$852	$8,254	$1,225	$11,876
	10%	786,103	$852	$1,225	$1,703	$16,508	$2,451	$23,753
OR	1%	25,686	$9	$14	$19	$2,158	$28	$3,201
	5%	128,430	$46	$69	$93	$10,788	$138	$16,006
	10%	256,860	$93	$138	$186	$21,576	$275	$32,012
PA	1%	87,223	$31	$46	$63	$7,327	$91	$10,588
	5%	436,114	$157	$228	$315	$36,634	$455	$52,939
	10%	872,228	$315	$455	$630	$73,267	$910	$105,877
SC	1%	27,223	$10	$14	$49	$1,143	$71	$1,657
	5%	136,114	$49	$71	$246	$5,717	$356	$8,285
	10%	272,229	$98	$142	$492	$11,434	$712	$16,570
UT	1%	15,584	$6	$8	$28	$655	$42	$981
	5%	77,921	$28	$42	$141	$3,273	$211	$4,907
	10%	155,842	$56	$84	$281	$6,545	$422	$9,813
WA	1%	42,703	$15	$22	$77	$1,794	$110	$2,566
	5%	42,703	$77	$110	$386	$8,968	$552	$12,829
	10%	42,703	$154	$221	$771	$17,935	$1,103	$25,658

Figure 31. Relative costs of using a network of human helpers to defeat CAPTCHAs while changing 1, 5, and 10 percent of the voters at the 11 websites having CAPTCHAs. Center right columns compare costs based on using named sources for data (Figure 27); the rightmost columns show costs based on using darknet sources for data (Figure 30).

State/ 1% of Voters		Total Costs with Automated CAPTCHA				Total Costs with Human Networked CAPTCHA			
		Named Sources		Darknet Sources		Named Sources		Darknet Sources	
AL	33,308	$2,061	8.1%	$90	0.9%	$2,061	4.6%	$90	0.2%
AK	505	$21	0.1%	$1	0.0%	$21	0.0%	$1	0.0%
AZ	35,885	$2,066	8.1%	$21	0.2%	$2,066	4.6%	$21	0.1%
CA	170,283	$342	1.3%	$580	5.8%	$342	0.8%	$580	1.5%
CO	35,711	$65	0.3%	$19	0.2%	$65	0.1%	$19	0.0%
CT	21,154	$46	0.2%	$73	0.7%	$46	0.1%	$73	0.2%
DE	6,820	$5	0.0%	$7	0.1%	$301	0.7%	$451	1.2%

State/ 1% of Voters		Total Costs with Automated CAPTCHA				Total Costs with Human Networked CAPTCHA			
		Named Sources		Darknet Sources		Named Sources		Darknet Sources	
GA	46,957	$336	1.3%	$134	1.3%	$336	0.7%	$134	0.3%
HI	7,069	$513	2.0%	$26	0.3%	$513	1.1%	$26	0.1%
IL	80,451	$1,452	5.7%	$1,020	10.1%	$1,452	3.2%	$1,020	2.6%
IN	48,292	$2,106	8.3%	$160	1.6%	$4,204	9.3%	$3,334	8.5%
KS	16,446	$230	0.9%	$45	0.4%	$230	0.5%	$45	0.1%
KY	31,053	$507	2.0%	$17	0.2%	$507	1.1%	$17	0.0%
LA	30,318	$2,056	8.1%	$79	0.8%	$2,056	4.5%	$79	0.2%
MD	39,001	$200	0.8%	$21	0.2%	$200	0.4%	$21	0.1%
MA	42,718	$78	0.3%	$115	1.1%	$78	0.2%	$115	0.3%
MI	74,546	$137	0.5%	$38	0.4%	$137	0.3%	$38	0.1%
MN	32,693	$106	0.4%	$82	0.8%	$106	0.2%	$82	0.2%
MO	41,909	$177	0.7%	$21	0.2%	$177	0.4%	$21	0.1%
NE	11,407	$525	2.1%	$11	0.1%	$1,487	3.3%	$1,354	3.5%
NV	16,864	$537	2.1%	$20	0.2%	$1,960	4.3%	$2,337	6.0%
NJ	58,193	$109	0.4%	$35	0.3%	$109	0.2%	$35	0.1%
NM	11,980	$2,022	8.0%	$37	0.4%	$2,022	4.5%	$37	0.1%
NY	124,933	$1,478	5.8%	$1,638	16.2%	$1,478	3.3%	$1,638	4.2%
OH	78,610	$229	0.9%	$330	3.3%	$1,965	4.3%	$2,828	7.2%
OR	25,686	$548	2.2%	$71	0.7%	$2,715	6.0%	$3,286	8.4%
PA	87,223	$211	0.8%	$277	2.7%	$7,570	16.8%	$10,910	27.9%
RI	7,709	$14	0.1%	$22	0.2%	$14	0.0%	$22	0.1%
SC	27,223	$492	1.9%	$359	3.6%	$1,675	3.7%	$2,073	5.3%
TX	151,011	$3,059	12.0%	$2,947	30.3%	$3,059	6.8%	$3,059	7.8%
UT	15,584	$34	2.1%	$51	0.5%	$711	2.7%	$1,066	2.7%
VT	4,698	$9	0.0%	$12	0.1%	$9	0.0%	$12	0.0%
VA	55,297	$2,101	8.3%	$29	0.3%	$2,101	4.7%	$29	0.1%
WA	42,703	$521	2.0%	$561	5.6%	$2,376	5.3%	$3,215	8.2%
DC	4,798	$9	0.0%	$12	0.1%	$9	0.0%	$12	0.0%
WV	12,749	$523	2.1%	$7	0.1%	$523	1.2%	$7	0.0%
National Dataset				$1,002				$1,002	
Totals	**1,531,787**	**$24,926**		**$10,081**		**$44,681**		**$39,087**	

Figure 32. Comparison of total attack costs for changing 1 percent of voter records on state websites using automated means to defeat CAPTCHAs versus using a semi-automated means with a network of human helpers. On the left are relative costs for using automated means to defeat CAPTCHAs with named sources for data, copied from Figure 27, and with darknet sources for data, copied from Figure 30. The totals on the right show costs when a network of humans is used to defeat CAPTCHAs instead of automated means. The 11 state websites that had CAPTCHAs were: CT, DE, IN, NE, NV, OH, OR, PA, SC, UT, and WA.

Results for Step 6e. Comparison of Total Attack Costs by State

In this sub-section, we rank states by total attack costs. Throughout these results, we base our assessments on two primary attack models that differ by the source of personal data used. In one model, the attacker acquires data through named government offices, data brokers or websites published on search engines. In the other model, the attacker uses a named data broker, a voter list from Texas, and two datasets from darknet sources (National Dataset).

State rankings varied by model. Figure 33 ranks states by the total attack costs for changing 1 percent of all voter records at the 36 state websites using data acquired from named sources. Texas is the most expensive at $3,059, followed by Indiana ($2,106), Virginia ($2,101), and Arizona ($2,066). Delaware is the least expensive at $5, followed by Vermont and the District of Columbia ($9), Rhode Island ($14), and Alaska ($21).

Figure 34 ranks the states by total attack costs for making the same voter records changes using the second model, the one that involves data acquired from darknet sources. Texas is again the most expensive at $3,059, followed by New York ($1,638), Illinois ($1,020), and then California ($580). The least expensive is Alaska at $1, followed by West Virginia ($7), Delaware ($7), and Nebraska ($11), the District of Columbia ($12), and then Vermont ($12).

Factors that determine the attack costs for states in these models are the number of voters in the state, the lowest price found for the personal data and government-issued numbers of the voters, and whether the website used CAPTCHAs.

Website editing of voter rolls, scraping data from data broker websites, and responding to CAPTCHAs are all per person costs. The more voters or attempts to impersonate voters, the greater these costs. States with more voters will therefore have greater attack costs for these activities, if all other factors remain the same.

Another factor that influences state rankings in the two models is reliance on SSNs. The darknet sources made the cost of acquiring SSNs insignificant, so it lowered the total attack costs of those states whose websites relied on SSNs alone.

Texas ranks as the most expensive in both models because its website required a voter ID number that was only available from its voter list, and we did not find a statewide Texas voter list on the darknet. We believe that the attacker would have no choice but to purchase the Texas voter list. We found a partial Texas voter list on the darknet for $15.98 (Figure 18) but not a statewide list. If other state websites required voter ID numbers also, then the attacker would have to acquire their voter lists too, which would drive the maximum attack costs to be those in the first model (Figure 35). On the other hand, if those same voter lists were available on the darknet at prices comparable to what we found for statewide voter lists from darknet sources (e.g., $30.97 for the Nevada voter list; see Figure 18), then the maximum attack costs would be about the same as those we recorded for the second model (Figure 36).

Put another way, if a statewide Texas voter list is available on the darknet, then its attack costs will plummet.

In short, we found two attack models that exploit the state websites based on different sources of data available at the time.

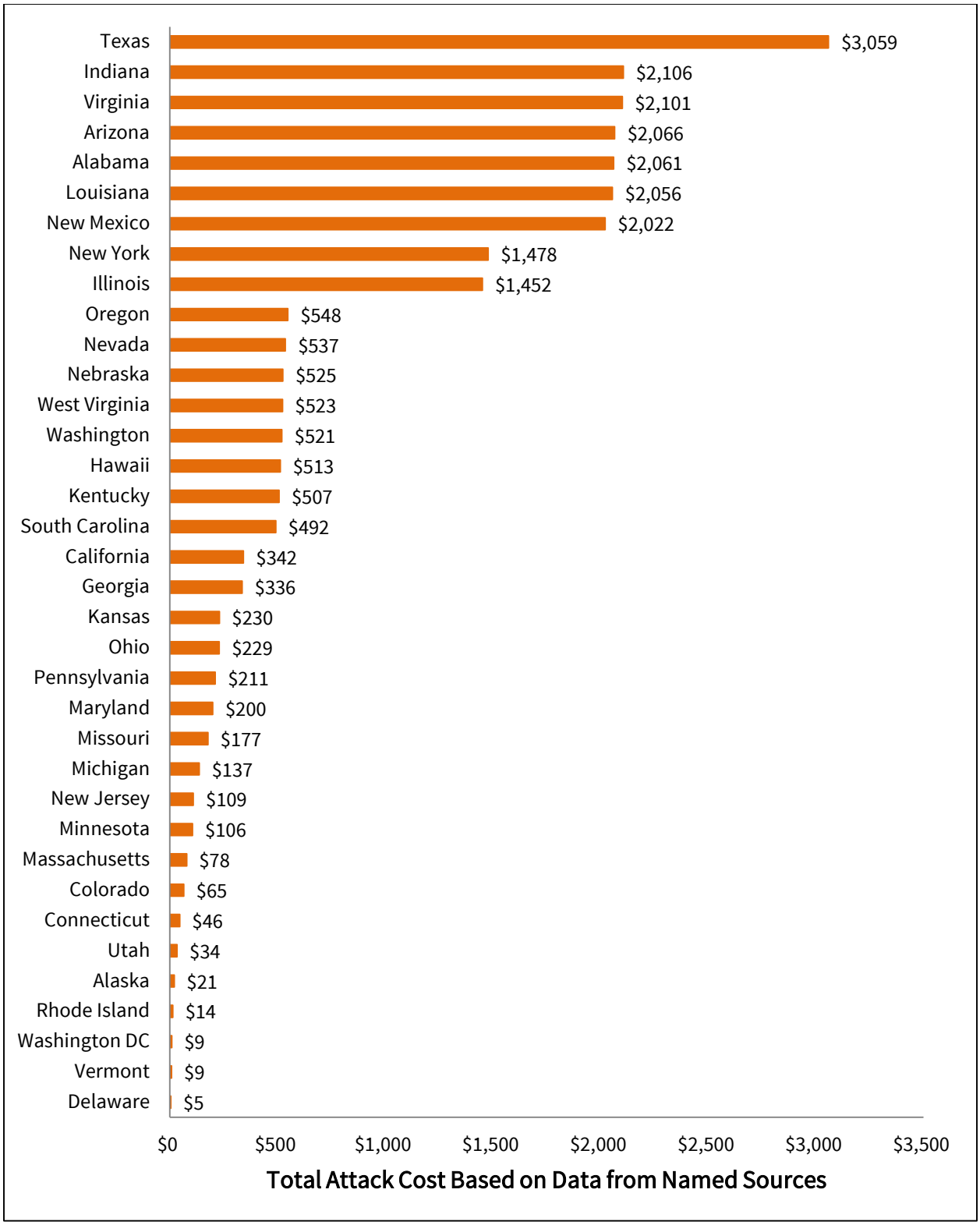


Figure 33. Comparison of total attack costs by state for an attack using data from named sources—i.e., data from government offices, data brokers, and websites published on search engines. See Figure 27 for details.

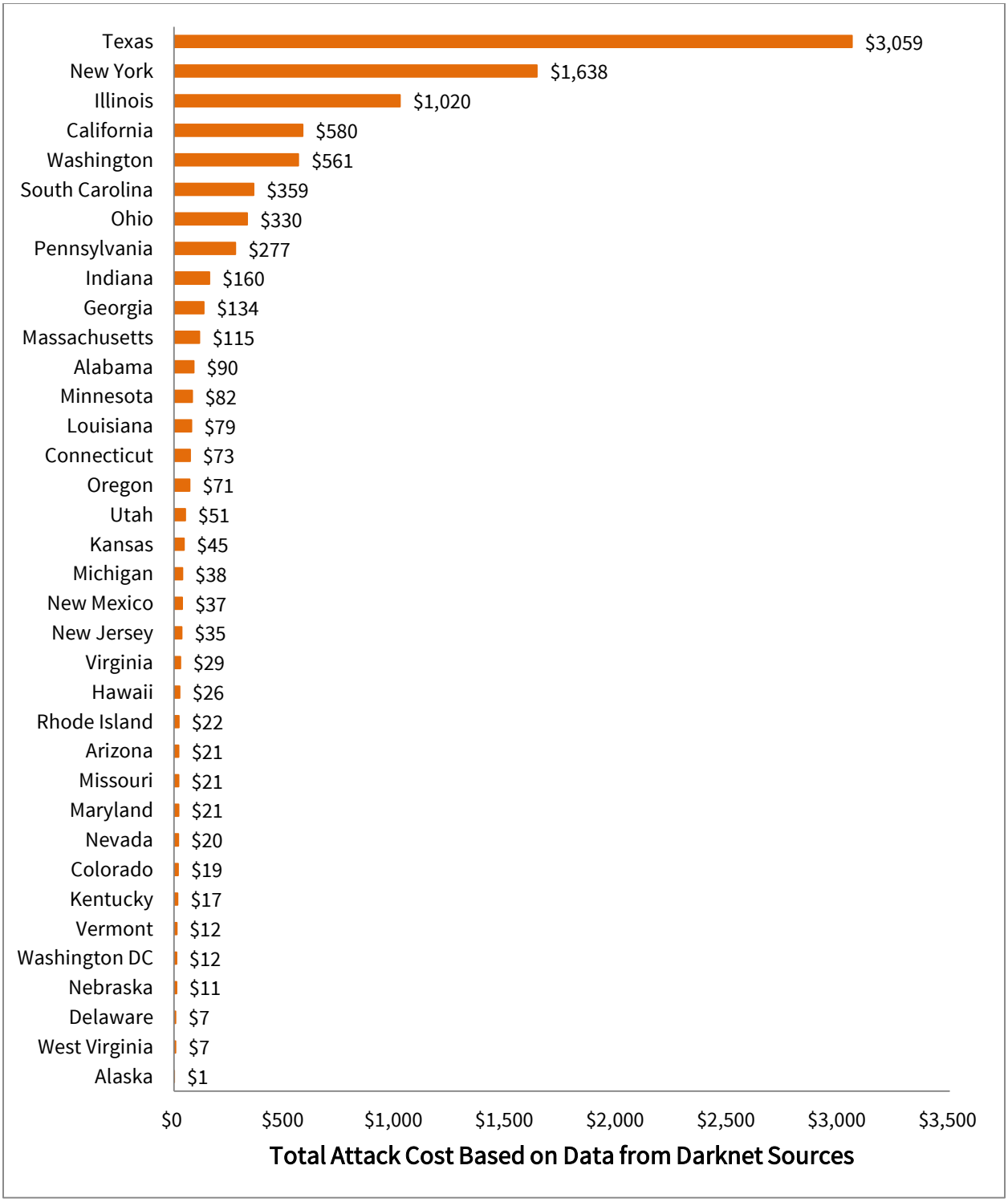


Figure 34. Comparison of total attack costs by state for an attack using data from a combination of two datasets from the darknet to jointly produce a National Dataset having the names, demographics, and SSNs of most adult Americans, access to a named data broker's website, and a copy of the Texas Voter List purchased from Texas. See Figure 30 for details.

Results: Closing thoughts on data and attack costs

We are not stating that our reported attack totals are the actual costs any attacker would have paid because specifics matter. For example, using a data broker that charges $1 per driver's license issue date is 100 times more expensive than using a data broker that charges $0.01 per lookup. Still, the models we constructed can be used to estimate total costs based on actual attack decisions made. For example, using a data broker that charges $1 per driver's license data would have had no impact in the 31 states that did not use driver's license issue dates. However, attempting to change 1 percent of the voters on all 36 websites and paying $1 instead of $0.01 for each driver's license issue date lookup would have been significantly more expensive, with the cost increasing from $10,081 to $432,026 using the darknet data option and from $24,926 to $446,983 using data from named sources.

Different kinds of data were more expensive to obtain, so some states could increase attack costs by adding additional requirements for website visitors to satisfy. This adds more data hurdles for online users, but would it be effective? Texas' website required a voter ID, which was only available on its voter list. Driver's license issuance dates required a per person lookup fee because the data was most widely available from a data broker through a search service. Document numbers on driver's licenses were more obscure but still available or inferable. CAPTCHAs increased the computational costs. The attack cost to submit address changes to 1 percent the voters on all 36 websites, assuming each website had all these requirements, goes from $24,926 to only $36,358. This suggests that field selection alone is an insufficient guard against voter identity theft attacks.

We showed that automated programs could defeat the kinds of CAPTCHAs we found on the 11 websites that had CAPTCHAs. As a result, total attack costs were low for those websites. By March 2017, after the presidential election, Google updated its reCAPTCHA v2 service to thwart the kind of audio-to-text automation we describe as capable of defeating those CAPTCHAs. Semi-automated attacks on CAPTCHAs still remain possible, but they are much more expensive than fully automated ones because the attacker must pay humans. We estimated that the cost of changing 1 percent of the voters using data from named sources increases from $24,926 to $44,681 when automated CAPTCHA responses are replaced with semi-automated ones. Even with the improvement Google made to its service, there is no guarantee that there does not exist another automation capable of defeating the new version and driving costs down again.

An attacker does not even need voter data to perpetrate an attack. We showed that using 2 datasets available on the darknet along with a data broker's website, an attacker could implement an attack for about $10,000 to change the addresses of 1 percent of the voters on 35 of the 36 websites. These 2 datasets, together costing $1,002, included the names, addresses, demographics, and SSNs of most adult Americans. An attacker can use this data, without SSNs, as a replacement for voter data in the attacks we described. Doing so requires making more attempts at a state website because the website will accept only some as

registered voters. As we reported, using general population data rather than a voter list increases the number of attempts made by an attacker by about 60 percent. A voter list usually includes voting history and party designation too, making it easier for an attacker to target specific kinds of voters to disenfranchise. However, an attacker can target specific kinds of voters from population data by using address and even race and gender, inferred by name, as the bases for targeting. More generally, consumer lists and club memberships could also be used to target voters presumed to be in favor of or opposed to a position, and can be used without a voter list.

Impersonating voters on the Texas website required voter ID numbers, so the voter list for Texas was necessary for acquiring needed voter ID numbers. An attacker could purchase the Texas voter list from the state for $1,160 to acquire voter date including the ID numbers. The partial voter list available on the darknet for $15.98 that included 4 percent of Texas voters also included voter ID numbers. The lesson from the Texas example is that using voter ID numbers on a website can force an attacker to have a voter list, but the availability of most voter lists at practical costs, as discussed above, suggests that the requirement of voter ID numbers on websites offers little or no protection when voter ID appears on the voter list or is inferable from other information.

Similarly, we found driver's license data readily available from data brokers or on darknet markets for all states.

Americans and American financial systems consider Social Security numbers (SSNs) to be personally identifying privately held information, but we showed that SSNs are widely available through named data brokers and on the darknet. Some brokers and darknet sources charge $10 per lookup for a targeted person or for a person matching a particular criterion. But in bulk, SSNs are widely available and inexpensive to acquire from named sources as well as on darknet markets. Most notably, we reported multiple offers on the darknet for bulk access to SSNs, including a dataset for $500 that reportedly included the SSNs of virtually all adult Americans. Restricting the search to named sources, we reported data broker websites providing unlimited searches for flat monthly fees from $40 to $60 per month. Ironically, because SSNs are so widely associated with financial, tax, and other records, SSNs are widely available. Twenty-one of the 36 websites required SSNs.

We found data brokers that provide driver's license data too, although fewer data brokers provided driver's license data than supplied SSNs. A data broker that provided SSNs and driver's license data tended to do so at the same bundled price. That is, the price for either SSN or driver's license information tended to be the same as acquiring both driver's license data and SSNs together. We also found this pattern on darknet offers, although we did not find a comprehensive dataset of driver's license data for sale on the darknet markets as we did for SSNs.

Some of the 33 websites that require driver's license data had driver's license numbers that could be acquired online. For example, one state had a website where, if the attacker provided the voter's SSN, the website would provide the voter's driver's license number. And 6 states had driver's license numbers that could be computed from demographics. No data purchase was needed for these states, although though two of them also required driver's license issuance dates, which necessitated using a data broker except when the issuance dates used by the state relate to the voter's birthday. Because driver's license numbers are not as widely associated with other kinds of records as are SSNs, the secondary availability of driver's license data is less than that of SSNs, but the numbers are still available at reasonable costs for an attacker.

An attack to change 1, 5, or 10 percent of the voter registrations on the 36 websites requires a lot of computation. We explained that an attacker could deploy 57 virtual cloud machines for one month at a cost of less than $1,000 to do the work needed to change 1 percent of the voter records at all 36 websites. That includes the machine time needed to defeat any CAPTCHAs and scrape information from a data broker website. Virtual machines available from multiple cloud services can be utilized on demand.

Architecting a scaled attack takes programming skill along the lines of a reasonably proficient web programmer. But people with little or no programming experience can modify existing code for the specifics of each website. We gave examples of the kinds of simple Python programs needed to do accomplish a similar task at a state website. We had students and summer interns, most of whom were newcomers to programming, revise and customize sample code to state website specifics to learn polling places. And of course, changing hundreds of voter records could be done manually, without any programming whatsoever. Automation becomes necessary at scale, if the goal is to change thousands or millions of voter records.

Discussion

Discussion: Summary of Findings

We showed that an attacker could impersonate voters and submit address changes that if accepted, could disenfranchise significant numbers of voters or disrupt elections. Automated attacks could have been perpetrated at scale, submitting changes to the voter records of 1, 2, 5, or even 10 percent of the records at all 36 websites by practical means, or by using limited resources to concentrate on specific regions or states. While our analysis focused on scenarios in which an attacker submitted voter address changes, similar approach and analysis holds for an attacker requesting absentee ballots, and in some states, submitting removals from the voter rolls.

The critical information needed consisted of the names, addresses, dates of birth, Social Security Numbers (SSNs) and driver's license data of voters, though no one state website requested all this information. As examples, Delaware allowed a voter to enter only {name,

date of birth, ZIP (postal code)}, Kentucky accepted {name, date of birth, SSN}, and Nevada required {name, date of birth, address, last 4 digits of SSN, driver's license number}.

An attacker could acquire needed voter names, demographics and government-issued numbers from government offices, data brokers, the deep web, or darknet markets. The total cost of an attack varied by the information requested at the website, the number of attempts needed to impersonate voters, the data sources used, CAPTCHAs, and the size of the voter population in the geographical area of interest. All CAPTCHAs found on the websites in 2016 could be defeated by automated means.

We surveyed data and computing expense to model the relative costs of different attack scenarios. We found that an attacker could purchase 2 datasets from the darknet, a voter list, and a subscription to a data broker's website to change 1 percent of the voter records on all 36 websites for a total attack cost of $10,081. We also reported that an attacker who wanted to avoid darknet purchases altogether could purchase needed data from government offices, data brokers, and publicized websites to change 1 percent of the voter records on all 36 websites for $24,926. Targeted areas were less, such as $1 for Alaska or $1,020 for Illinois.

Discussion: State Defenses

Attacks could be launched at these 36 websites, but would attacks be successful? The answer depends on back end office processes and state election practices. There are four functional stages to consider: (a) submission of change requests; (b) acceptance of changes; (c) database intelligence; and, (d) state election practices. See Figure 35.

The attack occurs when an imposter submits voter record changes on the state website; see Figure 35(a). We showed this is possible on all 36 websites. So, the first level of defense is any behind the scenes review of submitted change requests; see Figure 35(b). Would someone at a state election office notice or intervene? The second level of defense is the ability to surgically identify and undo specific database changes if an attack becomes subsequently known; see Figure 35(c). Does the database store meta information about voter roll changes? The final level of defense depends on state election practices that may limit the impact of an attack on the vote count; see Figure 35(d). Would the voter be able to cast a fully counted ballot even if the voter roll appeared chaotic? We asked state elections officials in August 2017 and found the following.

Some states maintain and store website access and change logs. These allow authorities to assess the nature and extent of problems if they surface and are useful for identifying patterns of IP addresses that connect with the website. Arizona, California, Indiana, Louisiana, Minnesota, New Jersey, Utah, Vermont, Washington, and West Virginia asserted to us that they do this. Others may too.

Many states review logs of change requests to determine usual activity such as an unexpected increase in the number of online registration changes; changes made in areas

likely to be close or pivotal in the election result; and the appearance of unusual or foreign IP addresses. Arizona, California, Louisiana, Minnesota, New Jersey, Utah, Vermont, Washington, and West Virginia asserted to us that they do this. Others may too.

Several states maintain internet and change logs for a period to include at least 3 months post-election in case voter complaints about unauthorized online changes to voter registration information need to be investigated. Arizona, California, Louisiana, Minnesota, New Jersey, Rhode Island, Utah, Vermont, Washington, and West Virginia asserted to us that they do this. Others may too.

Numerous states log database changes over time so that reversing changes is possible. Arizona, California, Louisiana, Minnesota, New Jersey, Rhode Island, Utah, Vermont, Washington, and West Virginia asserted to us that they do this. Others may too.

Less states reported having at least two fields in the voter record database to store the date on which the address was entered and to describe the means of the change (e.g., web change or postal service update). Arizona, California, Louisiana, Minnesota, New Jersey, Rhode Island, Utah, Washington, and West Virginia asserted to us that they do this. Others may too.

The United States Postal Service sends a confirmation to the old and new addresses when it receives an online change of address request [153]. Similarly, a few states mail a postcard to the old and/or new address of any voter whose address has been changed online to notify the voter that a change occurred. California, Indiana, New Jersey, Utah, Washington, and West Virginia asserted to us that they do this. Others may too. Indiana sends a mail notification of the address change to the new address and does not change the voter record unless confirmation is returned by mail within 7 days.

Some states allow contested voters the option of a countable provisional ballot, have same day registration, or allow the voter to sign an affidavit and cast a regular ballot. Arizona, California, Colorado, Indiana, Louisiana, Minnesota, New Jersey, Rhode Island, Utah, Vermont, Washington, and West Virginia asserted to us that they do this. Others may too.

Numerous states provide public service announcements for voters to check their registered voting information online regularly, similar to the advertisements encouraging credit report checking, and provide a way for voters to report problems. Louisiana, Minnesota, Rhode Island, Utah, Vermont, Washington, and West Virginia asserted to us that they do this. Others may too.

Many states report capturing logs for years, but a complication to the Riverside County incident last year was a lack of the necessary logs required to conduct the investigation. An absence of logs often hinders investigation and makes it difficult to refute accusations. We acquired copies of official letters between the Riverside County District Attorney and the California's Secretary of State and voter complaints and spoke with relevant parties regarding the matter. There is a factual basis for the allegations but the absence of sufficient

logs makes it difficult to completely explain the facts. After the 2016 incident, the California system began keeping more robust logs in 2017. For this reason, the Riverside County, California incident provides a teachable moment about the importance of logs and, the ability for election systems to improve when problems surface. Assessing problems, as this paper does, is critically important to continuing this cycle towards ever stronger election systems.

Even if every contested voter is allowed to cast a regular ballot, changing voter records at scale can make an election seem chaotic at precincts and lead to a loss of trust. Provisional ballots may appease voters at the polling place, but it is prudent to protect the integrity of the voter rolls to avoid them.

A couple of state election officials mentioned that their systems limit the number of accesses a single IP address may make to the website in a period of time. This seems problematic. A community center's voter registration drive may be wrongly halted after the first few registrations because its computers use the same IP address for registering lots of voters. On the other hand, an imposter could use a VPN network (e.g., NordVPN or IP Vanish [154]) to mount the attack where few submissions share the same IP address.

Election officials in most states, including in Indiana, Minnesota, Vermont, and Washington, mentioned a standard practice by which local election offices receive online change requests remotely by computer and have a human make the final decision to accept or reject. A Delaware election official reported that years ago this human-in-the-loop system noticed an inordinate number of paper applications and that some were fraudulent and then worked to help law enforcement bring criminal charges. A human in the loop may detect an unusually high number of changes but may not be able to detect fraudulent changes if the attacker submits small numbers of changes interspersed over time. As a next round of improvements, computer programs might offer a good aide to help detect unusual activity at the website and among change requests, even if small numbers over time. These kinds of improvements make work best with online processes.

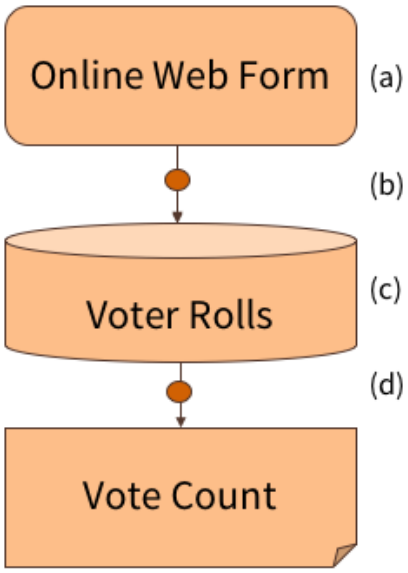


Figure 35. Stages of defenses for a voter identity theft: (a) online web submission; (b) acceptance of the submission; (c) voter rolls database; and, (d) election practices.

Discussion: Possible Attackers

Political motivations and financial incentives exist to exploit the kinds of attacks we describe. As examples, here are 3 kinds of potential attackers. (1) Criminal organizations have historically used their resources –money, violence, and loyalty—to influence democratic processes, and increasingly, criminal enterprises are becoming active online [155]. Criminal organizations might attempt voter identity theft attacks to exert control over or influence local elections. (2) Political data analytic companies have emerged as independent contractors with the technical know-how to manage, use and manipulate data to a campaign's advantage. Results from recent elections have been close in some states, within a few percentage of the vote cast [156]. Imagine a new operation that promises to shave percentage points on elections, having as their business secret the attacks we describe. (3) State actors have previously interfered in other countries' elections and effectively engaged in cyberwarfare activities [157]. A state actor could launch voter identity theft attacks to scramble voter records, thereby generating chaos and disturbance on election day, and undermining trust in election results. Or, a state actor might strategically target voters in an attempt to influence an election's outcome.

Historically, individuals and groups seeking to influence the outcome of a national or state election spend hundreds of millions of dollars to do so. The attacks we describe would cost far less, hundreds of dollars for most states, and less still for local or district elections.

Discussion: What We Did Not Show

We documented how voter identity theft attacks could have been launched at 36 state websites in terms of what personal information was needed; how an attacker might acquire needed information and defeat CAPTCHAs; and the ability and costs of automating requests to change voter rolls.

Our results do not claim imposters changed or deleted voter registrations or requested absentee ballots in 2016, or even attempted to do so. Instead, our results describe what an attacker could know and do to submit address changes to voter registrations in 2016.

The incident that Riverside County, California, experienced during the Republican primary in June 2016 seems to have been the first reported case of widespread voter identity theft using a voter records website.

We did not investigate whether anyone actually exploited the vulnerability we documented, so we are not asserting that the vulnerability had an actual impact on the 2016 election. We have shown that the vulnerability existed and that the attack could be done using practical resources to have a scaled impact.

We found no other technical problems with online voter registration systems, only with the ability to change addresses of registered voters online, whether newly registered online or not. Some of the websites were not even voter registration websites but motor vehicle or other websites where voter records could be changed.

The voter identity attacks we describe do not involve breaking into a computer or breaching the state's voter roll, as are often the concern of computer security. The described attacks would not account for additional votes being cast, as is often the concern of "voter fraud"; rather, the attacks we describe might account for fewer votes being counted and more provisional ballots being used.

Discussion: Online Versus Offline

The ability to fraudulently submit address changes to voter records is not an online problem. Almost every state election system accepts change of address forms by mail, even states that do not have online websites for changing voter information. For convenience, the U.S. Election Assistance Commission provides a single form for registering voters and changing addresses by mail that is accepted by all 35 states and the District of Columbia [158].

The single mail-in form requires name, address, date of birth, and either a driver's license number or all or part of an SSN. Two states require a race selection from among 6 choices be completed on the form. None of the states require driver's license issue date or document numbers. So, an imposter acquires the same data as for the online attack, without the additional per record cost for issuance dates. Instead of online submission, the imposter

must print completed forms with signatures imaged using cursive writing fonts, use a folder/inserter machine to prepare the printouts for mailing, and then physically mail the forms. Postage is $0.49 per change, which drives the total attack cost to change 1 percent of voters in all 35 states and the District of Columbia from $24,926 to $771,246, of which $750,576 (97 percent) is postal cost. It is not practical to launch an attack by mail on the scale possible online.

The ability to detect an attack increases too. Election staff would receive 1.5 million change requests in an attempt to change 1 percent of the voters in the 35 states and the District of Columbia. The increase in election office labor needed to type the information into the system would more likely be noticed even if distributed over thousands of counties. For these reasons, the vulnerability from attacks by mail appear more likely to occur on smaller scales, such as at a regional or local level.

Moving systems online did not introduce new problems, but changed the scale of existing problems. Voters should enjoy the benefits of online service without concern over voter identity theft and state offices should benefit from reduced expenses. Computers could help. Online systems could do more to detect attacks, such as computing whether an unusual number of changes are occurring and alert officials. These kinds of analytics underscore the need to continue to improve the security of our systems as we continue to improve our election services.

Discussion: Broader Contexts

Many news organizations have reported allegations of Russian meddling in the 2016 U.S. Presidential election (e.g., [159]). Most of the discourse about Russian hacking focuses on the possibility that hackers accessed email servers and unleashed propaganda campaigns. As we stated earlier, determining whether someone or some entity exploited the vulnerability we demonstrate is beyond the scope of this paper, but a state actor would have the resources needed to perpetrate the attack at scale.

In June 2017, Bloomberg News reported that Russian hacking of the 2016 U.S. election extended to 39 states, citing the possible compromise of voter registration and campaign finance files [160]. In Illinois, attempted changes reportedly included removing or altering voter data [160]. The article cites sources from the U.S. Department of Homeland Security but provides no specifics as to whether the particular attacks we describe were involved.

Beyond elections, the bigger question from our work is how does an American identify herself to a government website that only knows some basic information about her? Conversely, how does a government website authenticate Americans? Are American infrastructures –such as credit, tax refund, and elections– becoming too vulnerable to identity theft because of the availability of personal information on Americans? These are open important questions for computer security researchers, computer scientists, and technologists, and government agencies to solve. We present this study as a means to further study.

We were asked by several parties to consider not publishing this study. We decided to do so because whether we publish or not, does not make the possibility of these attacks go away. Instead, the secrecy would protect attackers that might exploit it and provide false confidence to election officials that past practices were sufficient. Then later if we learned that an attacker had been tweaking or disrupting elections along these lines, how irresponsible would we seem? The backlash might even ban or severely limit the future use of online voter registrations systems. Our goal in exposing problems and solutions is to help voters and state officials use new technologies with confidence.

Fair elections are critical to a healthy democracy. The ongoing debate over the possibility of voter fraud makes this more important than ever. We must rid these websites of the change-of-address vulnerability and pursue technologies that defeat voter identity theft to help assure the integrity of elections.

References

1. Horseman J. RIVERSIDE COUNTY: District attorney claims voters' party affiliations purposely changed without consent. July 8, 2016. http://www.pe.com/articles/county-807569-someone-information.html

2. Preal J. DA confirms hackers are culprit of voter registration changes. June 10, 2016. http://myvalleynews.com/most-relevant/voting-irregularities-disenfranchises-many-riverside-county/

3. Online Voter Registration: Trends in development and implementation. The Pew Charitable Trusts. May 2015. http://www.pewtrusts.org/~/media/assets/2015/05/ovr_2015_brief.pdf

4. Online Voter Registration. National Conference of Legislators. Accessed October 31, 2016. http://www.ncsl.org/research/elections-and-campaigns/electronic-or-online-voter-registration.aspx

5. State Online Voter Registration Systems. The Pew Charitable Trusts. April 11, 2016. http://www.pewtrusts.org/en/multimedia/data-visualizations/2016/state-online-voter-registration-systems

6. Understanding Online Voter Registration. The Pew Charitable Trusts. January 2014. http://www.pewtrusts.org/~/media/assets/2014/01/28/understanding_online_voter_registration.pdf?la=e

7. Barreto M, Glaser B, Mac Donald K, Collingwood L, Pedraza F, Pump B. Online Voter Registration (OLVR) Systems in Arizona and Washington: Evaluating Usage, Public Confidence and Implementation Processes. April 1, 2010.

https://www.sos.wa.gov/_assets/elections/Pew%20Report%20on%20Online%20Voter%20Registration%20in%20Arizona%20%26%20Washington%204-1-10.pdf

8. The American Voting Experience: Report and Recommendations of the Presidential Commission on Election Administration. January 2014. https://www.supportthevoter.gov/files/2014/01/Amer-Voting-Exper-final-draft-01-09-14-508.pdf

9. Shaw D, Ansolabhere S, Stewart C. A Brief Yet Practical Guide to Reforming U.S. Voter Registration Systems. Election Law Journal: Rules, Politics, and Policy. Vol. 14 Issue 1. March 2014. http://online.liebertpub.com/doi/abs/10.1089/elj.2014.0273

10. Burris A, Fishcher E. The Help America Vote Act and Election Administration: Overview and Selected Issues for the 2016 Election. Congressional Research Service. October 18, 2016. https://www.fas.org/sgp/crs/misc/RS20898.pdf

11. U.S. Government Accountability Office. Identity Theft and Tax Fraud: IRS Needs to Update Its Risk Assessment for the Taxpayer Protection Program. Report to Congressional Requesters. May 2016. http://www.gao.gov/assets/680/677406.pdf

12. Vaas L. IRS yanks anti-identity-theft tool IP PIN over 800 identity thefts. Naked Security. March 10, 2016. https://nakedsecurity.sophos.com/2016/03/10/irs-yanks-anti-identity-theft-tool-ip-pin-over-800-identity-thefts/

13. Vaas L. IRS hacked again – say goodbye to that PIN system!. Naked Security. June 27, 2016. https://nakedsecurity.sophos.com/2016/06/27/irs-hacked-again-say-goodbye-to-that-pin-system/

14. Munson L. Hackers compromise 100,000 IRS tax accounts with pre-stolen data. Naked Security. May 27, 2016. https://nakedsecurity.sophos.com/2015/05/27/hackers-compromise-100000-irs-tax-accounts-with-pre-stolen-data/

15. Harris S. Election Hackers Could Erase You. The Daily Beast. October 16, 2016. http://www.thedailybeast.com/articles/2016/10/16/election-hackers-could-erase-you.html

16. Absentee Ballot Rules. Vote.org. Accessed August 31, 2017. https://www.vote.org/absentee-voting-rules/ and an example from Alachua County, Florida at http://www.votealachua.com/MBRS.

17. Voting at a Polling Place after Applying to Vote by Mail. California Secretary of State. Accessed October 31, 2016. http://www.sos.ca.gov/elections/voting-resources/voting-california/if-you-applied/

18. § 24.2-712. Central absentee voter precincts; counting ballots. Virginia Law. Accessed October 31, 2016. http://law.lis.virginia.gov/vacode/title24.2/chapter7/section24.2-712/

19. Security Experts Question Maryland's Online Ballot System. CBS Baltimore. September 30, 2016. http://baltimore.cbslocal.com/2016/09/30/security-experts-question-marylands-online-ballot-system/

20. Greenspan R. Security experts question Maryland's online ballot system. Chicago Tribune. September 30, 2016. http://www.chicagotribune.com/news/sns-bc-md--cns-voting-system-20160930-story.html

21. Hicks J. Maryland moves forward with online voting despite warnings from cyber-experts. September 30, 2016. https://www.washingtonpost.com/local/md-politics/maryland-moves-forward-with-online-voting-despite-warnings-from-cyber-experts/2016/09/29/3de8f6a8-8644-11e6-a3ef-f35afb41797f_story.html

22. The National Voter Registration Act Of 1993 (NVRA). Department of Justice. September 1, 2016. https://www.justice.gov/crt/national-voter-registration-act-1993-nvra

23. Provisional Ballots. National Conference of State Legislators. June 19, 2015. http://www.ncsl.org/research/elections-and-campaigns/provisional-ballots.aspx#Fed%20Law

24. U.S. Election Assistance Commission. The Election Administration and Voting Survey: 2016 Comprehensive Report. https://www.eac.gov/assets/1/6/2016_EAVS_Comprehensive_Report.pdf

25. "If you move after the deadline to register to vote in a state election or primary, you should wait to update your registration until after the date of the election or primary, and return to vote at your previous polling place in Massachusetts." Elections: Voter Registration Information. Secretary of the Commonwealth of Massachusetts. Accessed October 31, 2016. https://www.sec.state.ma.us/ele/eleifv/howreg.htm

26. Absentee Ballot Deadlines. Vote.org. Accessed October 31, 2016. https://www.vote.org/absentee-ballot-deadlines/

27. Boutin P. The Secretive World of Selling Data About You. Newsweek. May 30, 2016. http://www.newsweek.com/secretive-world-selling-data-about-you-464789

28. Krebs B. Experian Sold Consumer Data to ID Theft Service. Krebs on Security. October 20, 2013. https://krebsonsecurity.com/2013/10/experian-sold-consumer-data-to-id-theft-service/

29. Consumer Data Information. Acxiom. Accessed October 31, 2016. http://www.acxiom.com/about-acxiom/privacy/consumer-data-information/

30. Krebs B. Experian Lapse Allowed ID Theft Service Access to 200 Million Consumer Records. Krebs on Security. March 10, 2014. http://krebsonsecurity.com/2014/03/experian-lapse-allowed-id-theft-service-to-access-200-million-consumer-records/

31. Federal Trade Commission. Data Broker Defendants Settle FTC Charges They Sold Sensitive Personal Information to Scammers. Press Release. February 18, 2016. https://www.ftc.gov/news-events/press-releases/2016/02/data-broker-defendants-settle-ftc-charges-they-sold-sensitive

32. American Institute of CPAs. AICPA Survey: One-in-four Americans Victimized by Information Security Breaches. Press Release. April 21, 2015. http://www.aicpa.org/press/pressreleases/2015/pages/aicpa-survey-one-in-four-americans-victimized-by-information-security-breaches.aspx

33. Identity Theft Resource Center. Identity Theft Resource Center Breach Report Hits Near Record High in 2015. January 25, 2016. http://www.idtheftcenter.org/Data-Breaches/2015databreaches.html

34. 21.5 Million Americans Affected By Government Data Breach. 10tv. July 10, 2015. http://www.10tv.com/article/215-million-americans-affected-government-data-breach

35. Herman B. Details of Anthem's massive cyberattack remain in the dark a year later. Modern Healthcare. March 30, 2016. http://www.modernhealthcare.com/article/20160330/NEWS/160339997

36. Krebs B. Experian Breach Affects 15 Million Consumers. Krebs on Security. October 2, 2015. https://krebsonsecurity.com/2015/10/experian-breach-affects-15-million-consumers/

37. Robertson J. Top Credit Agencies Say Hackers Stole Celebrity Reports. Bloomberg. March 12, 2013. https://www.bloomberg.com/news/articles/2013-03-12/equifax-transunion-say-hackers-stole-celebrity-reports

38. Reisinger D. Here's How Much Your Social Security Number Is Worth on the Dark Web. Fortune. August 3, 2016. http://fortune.com/2016/08/03/social-security-dark-web/

39. Krebs B. Credit Reports Sold for Cheap in the Underweb. Krebs on Security. March 13, 2013. https://krebsonsecurity.com/2013/03/credit-reports-sold-for-cheap-in-the-underweb/

40. Krebs B. Data Broker Giants Hacked by ID Theft Service. Krebs on Security. September 25, 2013. http://krebsonsecurity.com/2013/09/data-broker-giants-hacked-by-id-theft-service/

41. Bruer W, Perez E. Officials: Hackers breach election systems in Illinois, Arizona. CNN. August 30, 2016. http://www.cnn.com/2016/08/29/politics/hackers-breach-illinois-arizona-election-systems/

42. Bennet C, Geller E. FBI alert sparks fears that state voting systems are under digital assault. Politico. August 29, 2016. http://www.politico.com/story/2016/08/fbi-states-voting-systems-digital-assualt-227523

43. Greenberg A. Hack Brief: As FBI Warns Election Sites Got Hacked, All Eyes Are on Russia. Wired. August 29, 2016. https://www.wired.com/2016/08/hack-brief-fbi-warns-election-sites-got-hacked-eyes-russia/

44. Windrem R, Arkin W, Dilanian K. Russians Hacked Two U.S. Voter Databases, Officials Say. NBC News. August 30, 2016. http://www.nbcnews.com/news/us-news/russians-hacked-two-u-s-voter-databases-say-officials-n639551

45. Torres K. As Many As 7.5 Million Voter Records Involved in Georgia Data Breach. The Atlanta Journal-Constitution. March 3, 2017. http://www.myajc.com/news/state--regional-govt--politics/many-million-voter-records-involved-georgia-data-breach/rU2bMMc3tzGkuPvmbjlkgJ/

46. De Smet A. Unique ID – The numbers that control your life. Accessed October 31, 2016. http://www.highprogrammer.com/alan/numbers/index.html

47. De Smet A. US Driver's License Numbers - Maryland Michigan, Minnesota, perhaps others. Accessed October 31, 2016. http://www.highprogrammer.com/alan/numbers/dl_us_shared_mmm.html

48. Ohio Voter Files Download Page. State of Ohio. Accessed January 18, 2016. https://www6.sos.state.oh.us/ords/f?p=111:1

49. VoterListsOnline Voter Data Map. Aristotle. Accessed January 18, 2016. http://aristotle.com/vlomap/ and http://aristotle.com/data/datasolutions/ and VoterListsOnline.com

50. L2 Voter File. L2. Accessed January 18, 2016. http://www.l2political.com/products/data/

51. Fung B, Timberg C, and Gold M. A Republican Contractor's Database of Nearly Every Voter was Left Exposed on the Internet for 12 Days, Researcher says. The Washington Post. June 19, 2017. https://www.washingtonpost.com/news/the-switch/wp/2017/06/19/republican-contractor-database-every-voter-exposed-internet-12-days-researcher-says/?utm_term=.2703721db421

52. IRS. New IRS e-file Security and Privacy Standards FAQs. March 23, 2016. https://www.irs.gov/uac/new-irs-e-file-security-and-privacy-standards-faqs

53. Von Ahn L, Blum M, Hopper N, Langford J. CAPTCHA: using hard AI problems for security. Proceeding EUROCRYPT'03 Proceedings of the 22nd international conference on Theory and applications of cryptographic techniques. May 4, 2003. http://dl.acm.org/citation.cfm?id=1766196

54. Jeng A, Tseng C, Tseng D, Wang J. A Study of CAPTCHA and Its Application to User Authentication. Computational Collective Intelligence. Technologies and Applications. Volume 6422. 2010. http://link.springer.com/chapter/10.1007/978-3-642-16732-4_46

55. Yan J, Ahmad A. A low-cost attack on a Microsoft captcha. Proceedings of the 15th ACM conference on Computer and communications security. October 27, 2008. http://dl.acm.org/citation.cfm?id=1455839

56. National Institute of Standards and Technology. Guidelines on Securing Public Web Servers. Special Publication 800-44 Version 2. September 2007.http://csrc.nist.gov/publications/nistpubs/800-44-ver2/SP800-44v2.pdf

57. Zetter K. Wiseguys Plead Guilty in Ticketmaster CAPTCHA Case. Wired. November 19, 2010. https://www.wired.com/2010/11/wiseguys-plead-guilty/

58. Amazon Mechanical Turk. Working on Human Intelligence Tasks. Accessed January 18, 2016. https://www.mturk.com/mturk/welcome?variant=worker

59. Bursztein E, Aigrain J, and Moscicki A. The End is Nigh: Generic Solving of Text-Based CAPTCHAs. Usenix. August 4, 2014. https://www.usenix.org/system/files/conference/woot14/woot14-bursztein.pdf

60. Summers N. Vicarious Claims Its AI Software can Crack up to 90% of CAPTCHAs Offered by Google, Yahoo and Paypal. The Next Web. October 28, 2013. http://thenextweb.com/insider/2013/10/28/vicarious-claims-ai-software-can-now-crack-90-captchas-google-yahoo-paypal/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+TheNextWeb+(The+Next+Web+All+Stories)

61. Tung L. Google Algorithm Busts CAPTCHA with 99.8 Percent Accuracy. April 17, 2014. ZDNet. http://www.zdnet.com/article/google-algorithm-busts-captcha-with-99-8-percent-accuracy/

62. Lamo A. What is the Deep/Dark Web and How do You Access it? Quora. January 25, 2016. https://www.quora.com/What-is-the-deep-dark-web-and-how-do-you-access-it

63. Microsoft. How do I get a Skype Number? Skype. Accessed March 2017. https://support.skype.com/en/faq/FA256/how-do-i-get-a-skype-number

64. Tschabitscher H. Top Free Email Services 2017. March 2017. https://www.lifewire.com/top-free-email-services-1171481

65. Fox-Brewster T. 191 Million US Voter Registration Records Leaked in Mystery Database. Forbes. http://www.forbes.com/sites/thomasbrewster/2015/12/28/us-voter-database-leak/#132c225d1bb9

66. McDonald M, Licari P, and Merivaki L. U.S. Voter Information. United States Elections Project. 2015. http://voterlist.electproject.org/home

67. Colorado Voter List. Download Colorado Statewide Voer List. 2013, 2014, 2015, 2016, and 2017. http://coloradovoters.info/download.html

68. Connecticut Voter List. 2013, 2014, 2016, and 2017. http://connvoters.com/download.html

69. Delaware Voter List. 2012, 2014 and 2015. http://delawarevoters.info/downloads.html

70. Michigan Voter List. Download the Michigan Registered Voter List. 2013, 2014, 2015, 2016, and 2017. http://michiganvoters.info/download.html

71. Rhode Island Voter List. Registered Voters in the State of Rhode Island and Providence Plantations, USA. 2015 and 2017. http://rivoters.com/

72. Utah Voters List. 2014. http://web.archive.org/web/20140327211536/http://utvoters.com/

73. District of Columbia Voter List. DC BOEE Registered Voters Citywide. 2016. https://www.dcboee.org/pdf_files/ListOfRegisteredVoters051616.pdf

74. Catalist. https://www.catalist.us/

75. eMerges.com. http://www.emerges.com/

76. NationBuilder. http://nationbuilder.com/voterfile

77. Alabama Secretary of State. Alabama Votes. https://www.alabamavotes.gov/ and http://sos.alabama.gov/alabama-votes

78. State of Alaska, Division of Elections. Index of Resources. https://www.elections.alaska.gov/doc/forms/H23.pdf

79. Arizona Secretary of State. Voter Registration & Historical Election Data. http://voterlist.electproject.org/states/arizona

80. California Secretary of State. Elections and Voter Information. http://www.sos.ca.gov/elections/contact/

81. Colorado Secretary of State. Election Data. http://www.sos.state.co.us/pubs/info_center/fees/elections.html#ELECDATA

82. Connecticut Secretary of State. Centralized Voter Registration Information. http://www.ct.gov/sots/cwp/view.asp?a=3179&q=532994

83. State of Delaware. Freedom of Information Act. http://elections.delaware.gov/foia.shtml

84. Georgia Secretary of State. Order Voter Registration Lists and Files. http://sos.ga.gov/index.php/elections/order_voter_registration_lists_and_files

85. State of Hawaii Office of Elections. http://voterlist.electproject.org/states/hawaii and http://elections.hawaii.gov/

86. Illinois State Board of Elections. Computerized Registration Data. https://www.elections.il.gov/votinginformation/computerizedvoterdata.aspx

87. Indiana Elections Division. http://www.in.gov/sos/elections/ and http://voterlist.electproject.org/states/indiana

88. Kansas Secretary of State. Voter Registration Data Request. http://www.kssos.org/forms/elections/CVR.pdf

89. Kentucky State Board of Elections. Request for Voter Registration Data. http://elect.ky.gov/SiteCollectionDocuments/Request%20for%20Voter%20Registration%20Data/Request%20for%20Voter%20Registration%20Data%20Form.pdf

90. Louisiana Secretary of State. Voter List Charges and Other Information. http://www.sos.la.gov/ElectionsAndVoting/PublishedDocuments/VoterListChargesAndInfo.pdf

91. Maryland State Board of Elections. Voter Registration. http://www.elections.state.md.us/voter_registration/stats.html

92. Massachusetts Secretary of State. http://www.sec.state.ma.us/ele/eleifv/howreg.htm and http://voterlist.electproject.org/states/massachussets

93. Michigan Secretary of State. Elections in Michigan. http://www.michigan.gov/sos/0,1607,7-127-1633---,00.html and http://voterlist.electproject.org/states/michigan

94. Minnesota Secretary of State. http://www.sos.state.mn.us/ and http://voterlist.electproject.org/states/minnesota

95. Missouri Secretary of State. Elections and Voting. http://s1.sos.mo.gov/elections and http://voterlist.electproject.org/states/missouri

96. Nebraska Secretary of State Elections Division. Public Service Request Form, Nebraska Central Voter Registration System. http://www.sos.ne.gov/elec/pdf/voter-file-affidavit.pdf

97. Nevada Secretary of State. http://nvsos.gov/sos and http://voterlist.electproject.org/states/nevada

98. New Jersey Secretary of State. New Jersey Division of Elections. http://www.nj.gov/state/elections/ and http://voterlist.electproject.org/states/new-jersey

99. New Mexico Secretary of State. Voter Data Request. http://www.sos.state.nm.us/Elections_Data/Voter_Data_Request.aspx

100. New York State Board of Elections. Freedom of Information Requests (FOIL). Request for Access to Public Voter Registration Data. http://www.elections.ny.gov/FoilRequests.html

101. Oregon Secretary of State. Oregon Election History, Statistics and Data. http://sos.oregon.gov/elections/Pages/electionsstatistics.aspx and http://sos.oregon.gov/elections/Pages/request-for-voter-list.aspx

102. Commonwealth of Pennsylvania Department of State. PA Full Voter Export. https://www.pavoterservices.pa.gov/Pages/PurchasePAFULLVoterExport.aspx

103. Rhode Island Board of Elections. http://www.elections.state.ri.us/ and http://voterlist.electproject.org/states/rhode-island

104. South Carolina Election Commission. https://www.scvotes.org/ and http://voterlist.electproject.org/states/south-carolina

105. Texas Secretary of State. Abot the Elections Division. http://www.sos.state.tx.us/elections/ and http://voterlist.electproject.org/states/texas

106. Utah Lieutenant Governor Elections. Utah Voter Database. https://elections.utah.gov/voterdatabase

107. Vermont Secretary of State. https://www.sec.state.vt.us/elections.aspx and http://voterlist.electproject.org/states/vermont

108. Virginia Department of Elections. http://www.elections.virginia.gov/candidatepac-info/client-services/

109. Washington Secretary of State. Elections and Voting. https://www.sos.wa.gov/elections/vrdb/extract-requests.aspx

110. District of Columbia Board of Elections. Request for Voter Registration Data, Maps or Petition Signatures. https://www.dcboee.org/pdf_files/Data_Request_Form.pdf

111. West Virginia Secretary of State. Request for Voter Registration Data Instructions. http://www.sos.wv.gov/elections/forms/Documents/Forms%20-%20General/voter%20data%20request%20071213.pdf

112. Clark County Nevada Election Department. Voter List Data Files by District. http://www.clarkcountynv.gov/election/Pages/VoterDataFiles.aspx

113. U.S. Bureau of the Census. First names by Gender from the 1990 Census. Accessed March 2017. https://www.census.gov/topics/population/genealogy/data/1990_census/1990_census_namefiles.html

114. Dixon P. World Privacy Forum. Data Brokers Opt Out List. January 2017. https://www.worldprivacyforum.org/2013/12/data-brokers-opt-out/

115. Angwin J. Privacy Tools: Opting out from data brokers. Propublica. January 30, 2014 https://www.propublica.org/article/privacy-tools-opting-out-from-data-brokers

116. The Gramm-Leach-Bliley Act of 1999 (Financial Modernization Act of 1999) (15 U.S.C. 6801 to 6809). https://www.law.cornell.edu/uscode/text/15/6801

117. U.S. Social Security Administration. Social Security Number Randomization. https://www.ssa.gov/employer/randomization.html

118. Acquisti A and Gross R. Predicting Social Security numbers from public data. Proceedings of the National Academy of Sciences of the United States. May 5, 2009. http://www.pnas.org/content/106/27/10975.abstract

119. Corragroup. Motor Vehicle Records: 50 state guide to MVR input requirements. Accessed March 2017. http://www.corragroup.com/motor-vehicle-records.html

120. Nebraska Department of Motor Vehicles. Driver Privilege Services. Accessed March 2017. https://www.nebraska.gov/dmv/reinstatements/client.cgi Update August 2017: this website no longer displays a driver's license if one was not provided.

121. Russell R. Index (Soundex). U.S. Patent number 1,261,167 (filed Oct. 25, 1917). Issued 1923. # Re.15,582.

122. Repici D. Understandin the SoundEx Algorithm. CreativstSoftware. 2002. http://creativyst.com/Doc/Articles/SoundEx1/SoundEx1.htm

123. De Smet A. Driver's License Calculator: Illinois. 2013. http://www.highprogrammer.com/alan/numbers/dl_us_shared.html

124. De Smet A. Driver's License Calculator: Illinois. 2013. http://www.highprogrammer.com/cgi-bin/uniqueid/dl_il

125. De Smet A. U.S. Driver's License Numbers: Maryland, Michigan, Minnesota, perhaps others. 2003. http://www.highprogrammer.com/alan/numbers/dl_us_shared_mmm.html

126. De Smet A. Driver's License Calculator: Maryland. 2013. http://www.highprogrammer.com/cgi-bin/uniqueid/dl_md

127. De Smet A. Driver's License Calculator: Michigan. 2013. http://www.highprogrammer.com/cgi-bin/uniqueid/dl_mi

128. De Smet A. Driver's License Calculator: Minnesota. 2013. http://www.highprogrammer.com/cgi-bin/uniqueid/dl_mn

129. De Smet A. U.S. Driver's License Numbers: Florida, Illinois, Wisconis, perhaps others. 2002, 2013. http://www.highprogrammer.com/alan/numbers/dl_us_shared.html

130. De Smet A. New Jersey Driver's License Numbers. 2004. http://www.highprogrammer.com/alan/numbers/dl_us_nj.html

131. De Smet A. New York (Pre 1992) Driver's License Numbers. 2004. http://www.highprogrammer.com/alan/numbers/dl_us_ny92.html

132. De Smet A. Driver's License Calculator: New York. 2013. http://www.highprogrammer.com/cgi-bin/uniqueid/dl_ny92

133. De Smet A. Washington Driver's License Numbers. 2003. http://www.highprogrammer.com/alan/numbers/dl_us_wa.html

134. De Smet A. Driver's License Calculator: Washington. 2013. http://www.highprogrammer.com/cgi-bin/uniqueid/dl_wa

135. Alphabay Market Vendor Level/ Trust Level Explained. Reddit. https://www.reddit.com/r/AlphaBayMarket/comments/5qw3of/vendor_level_trust_level_explained/

136. Grams Search Engine. Darknet Tor address: http://grams7enufi7jmdl.onion/

137. Credit Card Glossary: Terms and Definitions. CreditCards.com http://www.creditcards.com/glossary/term-fullz.php

138. Alphabay Market. Accessed 2016. Closed 2017. Darknet Tor address: http://pwoah7foa6au2pul.onion/

139. Hansa Market. Accessed 2016. Closed 2017. Darknet Tor address: http://hansamkt2rr6nfg3.onion/

140. AlphaBay, the Largest Online 'Dark Market,' Shut Down. U.S. Department of Justice. Press Release. July 20, 2017. https://www.justice.gov/opa/pr/alphabay-largest-online-dark-market-shut-down

141. Dream Market. Darknet Tor address: http://lchudifyeqm4ldjj.onion/

142. dbWorld Database Market. Darknet Tor address: http://dbworldtheunglun.onion/

143. Isikoff M. Chinese hacked Obama, McCain Campaigns, took Internal Documents, Official Say. NBC News. June 6, 2013. http://investigations.nbcnews.com/_news/2013/06/06/18807056-chinese-hacked-obama-mccain-campaigns-took-internal-documents-officials-say

144. Miller Z. Obama, Romney Campaigns Subject to Repeated Hacking Attempts in 2012. Time. May 7, 2013. http://swampland.time.com/2013/05/07/obama-romney-campaigns-subject-to-repeated-hacking-attempts-in-2012/

145. Amazon Mechanical Turk Pricing. November 2016. https://requester.mturk.com/pricing

146. D'Onfro J. Amazon just increased prices on a service you've probably never heard of, and some researchers are really upset. Business Insider. June 23, 2015. http://www.businessinsider.com/amazon-mechanical-turk-price-changes

147. Perez S. Google Now Using ReCAPTCHA to Decode Street View Addresses. Tech Crunch. March 29, 2012. https://techcrunch.com/2012/03/29/google-now-using-recaptcha-to-decode-street-view-addresses/

148. Google reCAPTCHA. Frequently Asked Questions. https://developers.google.com/recaptcha/docs/faq

149. Cloud Speech API. Google. https://cloud.google.com/speech/

150. East-Ee. ReBreakCaptcha. February 28, 2017. https://github.com/eastee/rebreakcaptcha, https://east-ee.com/2017/02/28/rebreakcaptcha-breaking-googles-recaptcha-v2-using-google/

151. Microsoft Azure Virtual Machines. https://azure.microsoft.com/en-us/

152. Reported Voting and Registration, by Sex and Single Years o Age. U.S. Bureau of the Census. November 2016. https://www2.census.gov/programs-surveys/cps/tables/p20/580/table01.xls

153. Palmer K. Change of Address? USPS will Confirm. U.S. News and World Reports. February 27, 2008. https://money.usnews.com/money/blogs/alpha-consumer/2008/02/27/change-of-address-usps-will-confirm

154. Eddy M. The Best VPN Services of 2017. PC Magazine. August 28, 2017. https://www.pcmag.com/article2/0,2817,2403388,00.asp

155. Raab S. Five Families: The Rise, Decline, and Resurgence of America's Most Powerful Mafia Empires. St Martin's Press. Reprint 2016.

156. 2016 Presidential Election Results. Key President Election Results by State. Politico. December 13, 2016. http://www.politico.com/mapdata-2016/2016-election/results/map/president/

157. Strauss S. Cyberwar on Elections is Just Starting. USA Today. March 30, 2017. https://www.usatoday.com/story/opinion/2017/03/30/cyber-war-on-us-elections-just-begun-russia-election-hacking-steven-strauss-column/99783076/

158. National Mail Voter Registration Form. U.S. Election Assistance Commission. https://www.eac.gov/assets/1/6/Federal_Voter_Registration_8-31-17_ENG.pdf

159. Serhan Y. The Russian Hack of the DNC. The Atlantic. October 7, 2016. http://www.theatlantic.com/news/archive/2016/10/russia-hack-dnc/503405/

160. Riley M and Robertson J. Russian CyberHacks on U.S. Electoral System Far Wider Than Previously Known. Bloomberg News. June 13, 2017. https://www.bloomberg.com/news/articles/2017-06-13/russian-breach-of-39-states-threatens-future-u-s-elections

Authors

Latanya Sweeney is Professor of Government and Technology in Residence at Harvard University, X.D. and Nancy Yang Faculty Dean of Currier House at Harvard, Director of the Data Privacy Lab at Harvard, Editor-in-Chief of Technology Science, and was formerly the Chief Technology Officer of the U.S. Federal Trade Commission. She earned her PhD in computer science from the Massachusetts Institute of Technology and her undergraduate degree from Harvard. More information about Dr. Sweeney is available at her website at latanyasweeney.org. As Editor-In-Chief of *Technology Science*, Professor Sweeney was recused from the review of this paper.

Ji Su Yoo is an experienced researcher at Harvard University's Data Privacy Lab and at the Institute for Quantitative Social Science. She graduated from Harvard College with a B.A. in Social Studies. Currently, she is interested in the intersection between government and technology and how their interaction can impact privacy and data analysis in health care. She is also interested in exposing how technology and its use in government and policy-making may exacerbate or reflect existing political and societal inequalities. Ji Su hopes to attend graduate school and to continue conducting research that will inform public interest issues. As Managing Editor of *Technology Science*, Ji Su Yoo was recused from the review of this paper.

Jinyan Zang is a PhD student in Government at Harvard University and a 2017 recipient of the Paul and Daisy Soros Fellowships for New Americans. He is an experienced researcher on consumer protection, data security, and privacy issues as a Research Fellow at the Federal Trade Commission and a Research Analyst at Harvard University's Data Privacy Lab and at the Institute for Quantitative Social Science. He graduated cum laude in 2013 from Harvard College with a B.A. in Economics. As Managing Editor of *Technology Science*, Jinyan Zang was recused from the review of this paper.

The authors want to thank our 2016 Technology Science summer interns who identified some of the state websites; specifically, we thank Da'Von Boyd, Evelyn Chen, Asia DaCosta, and Rachel Matheson (listed alphabetically). We also acknowledge the automation support done by Natalie Altman, Evelyn Chen, and Anupama Jeevan (listed alphabetically). We also thank people for investigating their own information on their own state websites, including Raquel Hill and L. Jean Camp. We appreciate our communications with District Attorney Hestrin and California Secretary of State's Chief of Elections Jana Lean about the Riverside County, California investigation. We recognize the timely contributions made by state election officials who promptly responded to our outreach through phone calls, meetings and email communications and the facilitation of those by Leslie Reynolds at the National Association of Secretaries of State and Judd Choate at the National Association of State Election Directors. This work was funded in part by a grant from the Ford Foundation in support of the Data Privacy Lab at Harvard University, a grant from the John D. and Catherine T. MacArthur Foundation, and a grant from the Knight Foundation.

Referring Editor: Cheryl Brown

Citation

Sweeney L, Yoo J, Zang J. Voter Identity Theft: Submitting Changes to Voter Registrations Online to Disrupt Elections. Technology Science. 2017090601. Version 2. September 6, 2017. http://techscience.org/a/2017090601

Data

To be announced.

Appendix

Updated to Version 2 on 12/12/17

The revised version of this paper includes a small update to the language of the highlights. See Version 1 published on September 6, 2017

Below are supporting tables to compute driver's license numbers from demographics.

A.1. Illinois Driver License Encoding Table for First Name and Middle Initial [123]

Look up the first name in the following table.

Name	Code	Name	Code	Name	Code
Albert	20	Elizabeth	220	Margaret	560
Alice	20	Florence	260	Martin	560
Ann	40	Frank	260	Marvin	580
Anna	40	George	300	Mary	580
Anne	40	Grace	300	Melvin	600
Annie	40	Harold	340	Mildred	600
Arthur	40	Harriet	340	Patricia	680
Bernard	80	Harry	360	Paul	680
Bette	80	Hazel	360	Richard	740
Bettie	80	Helen	380	Robert	760
Betty	80	Henry	380	Ruby	740

Carl	120	James	440	Ruth	760
Catherine	120	Jane	440	Thelma	820
Charles	140	Jayne	440	Thomas	820
Clara	140	Jean	460	Walter	900
Donald	180	Joan	480	Wanda	900
Dorthy	180	John	460	William	920
Edward	220	Joseph	480	Wilma	920

If the name is not in the table above, then look up the first initial in the table below.

Initial	Code	Initial	Code	Initial	Code	Initial	Code
A	0	H	320	O	640	V	860
B	60	I	400	P	660	W	880
C	100	J	420	Q	700	X	940
D	160	K	500	R	720	Y	960
E	200	L	520	S	780	Z	980
F	240	M	540	T	800		
G	280	N	620	U	840		

Look up the middle initial in the following table.

Initial	Code	Initial	Code	Initial	Code	Initial	Code
A	1	H	8	O	14	V	18
B	2	I	9	P	15	W	19
C	3	J	10	Q	15	X	19
D	4	K	11	R	16	Y	19
E	5	L	12	S	17	Z	19
F	6	M	13	T	18		
G	7	N	14	U	18		

A.2. Maryland, Michigan, and Minnesota Driver License Encoding Tables [125]

Lookup the first name in the following table. If the exact name is not found, use the longest prefix that is on the table.

a	27	aa	28	ab	29	ac	30	ad	31
ae	32	af	33	ag	34	ah	35	ai	36
aj	37	ak	38	al	39	ala	40	alb	41
alc	42	ald	43	ale	44	alf	45	alg	46
alh	47	ali	48	alj	49	alk	50	all	51
alm	52	aln	53	alo	54	alp	55	alq	56
alr	57	als	58	alt	59	alu	60	alv	61
alw	62	alx	63	aly	64	alz	65	am	66

an	67	ao	68	ap	69	aq	70	ar	71
as	72	at	73	au	74	av	75	aw	76
ax	77	ay	78	az	79	b	80	ba	81
bb	82	bc	83	bd	84	be	85	bf	86
bg	87	bh	88	bi	89	bj	90	bk	91
bl	92	bm	93	bn	94	bo	95	bp	96
bq	97	br	98	bs	99	bt	100	bu	101
bv	102	bw	103	bx	104	by	105	bz	106
c	107	ca	108	cb	109	cc	110	cd	111
ce	112	cf	113	cg	114	ch	115	ci	116
cj	117	ck	118	cl	119	cm	120	cn	121
co	122	cp	123	cq	124	cr	125	cs	126
ct	127	cu	128	cv	129	cw	130	cx	131
cy	132	cz	133	d	134	da	135	db	136
dc	137	dd	138	de	139	df	140	dg	141
dh	142	di	143	dj	144	dk	145	dl	146
dm	147	dn	148	do	149	dp	150	dq	151
dr	152	ds	153	dt	154	du	155	dv	156
dw	157	dx	158	dy	159	dz	160	e	161
ea	162	eb	163	ec	164	ed	165	eda	166
edb	167	edc	168	edd	169	ede	170	edf	171
edg	172	edh	173	edi	174	edj	175	edk	176
edl	177	edm	178	edn	179	edo	180	edp	181
edq	182	edr	183	eds	184	edt	185	edu	186
edv	187	edw	188	edward	189	edx	190		
edy	191	edz	192	ee	193	ef	194	eg	195
eh	196	ei	197	ej	198	ek	199	el	200
ela	201	elb	202	elc	203	eld	204	ele	205
elf	206	elg	207	elh	208	eli	209	elizabeth	210
elj	211	elk	212	ell	213	ellen	214	elm	215
eln	216	elo	217	elp	218	elq	219	elr	220
els	221	elt	222	elu	223	elv	224	elw	225
elx	226	ely	227	elz	228	em	229	en	230
eo	231	ep	232	eq	233	er	234	es	235
et	236	eu	237	ev	238	ew	239	ex	240
ey	241	ez	242	f	243	fa	244	fb	245
fc	246	fd	247	fe	248	ff	249	fg	250
fh	251	fi	252	fj	253	fk	254	fl	255
fm	256	fn	257	fo	258	fp	259	fq	260
fr	261	fs	262	ft	263	fu	264	fv	265

fw	266	fx	267	fy	268	fz	269	g	270
ga	271	gb	272	gc	273	gd	274	ge	275
gf	276	gg	277	gh	278	gi	279	gj	280
gk	281	gl	282	gm	283	gn	284	go	285
gp	286	gq	287	gr	288	gs	289	gt	290
gu	291	gv	292	gw	293	gx	294	gy	295
gz	296	h	297	ha	298	hb	299	hc	300
hd	301	he	302	henry	303	hf	304	hg	305
hh	306	hi	307	hj	308	hk	309	hl	310
hm	311	hn	312	ho	313	hp	314	hq	315
hr	316	hs	317	ht	318	hu	319	hv	320
hw	321	hx	322	hy	323	hz	324	i	325
ia	326	ib	327	ic	328	id	329	ie	330
if	331	ig	332	ih	333	ii	334	ij	335
ik	336	il	337	im	338	in	339	io	340
ip	341	iq	342	ir	343	is	344	it	345
iu	346	iv	347	iw	348	ix	349	iy	350
iz	351	j	352	ja	353	jaa	354	jab	355
jac	356	jad	357	jae	358	jaf	359	jag	360
jah	361	jai	362	jaj	363	jak	364	jal	365
jam	366	james	367	jan	368	jao	369	jap	370
jaq	371	jar	372	jas	373	jat	374	jau	375
jav	376	jaw	377	jax	378	jay	379	jaz	380
jb	381	jc	382	jd	383	je	384	jea	385
jeb	386	jec	387	jed	388	jee	389	jef	390
jeg	391	jeh	392	jei	393	jej	394	jek	395
jel	396	jem	397	jen	398	jeo	399	jep	400
jeq	401	jer	402	jes	403	jet	404	jeu	405
jev	406	jew	407	jex	408	jey	409	jez	410
jf	411	jg	412	jh	413	ji	414	jj	415
jk	416	jl	417	jm	418	jn	419	jo	420
joa	421	job	422	joc	423	jod	424	joe	425
jof	426	jog	427	joh	428	john	429	joi	430
joj	431	jok	432	jol	433	jom	434	jon	435
joo	436	jop	437	joq	438	jor	439	jos	440
joseph	441	jot	442	jou	443	jov	444	jow	445
jox	446	joy	447	joz	448	jp	449	jq	450
jr	451	js	452	jt	453	ju	454	jv	455
jw	456	jx	457	jy	458	jz	459	k	460
ka	461	kb	462	kc	463	kd	464	ke	465

kf	466	kg	467	kh	468	ki	469	kj	470
kk	471	kl	472	km	473	kn	474	ko	475
kp	476	kq	477	kr	478	ks	479	kt	480
ku	481	kv	482	kw	483	kx	484	ky	485
kz	486	l	487	la	488	lb	489	lc	490
ld	491	le	492	lea	493	leb	494	lec	495
led	496	lee	497	lef	498	leg	499	leh	500
lei	501	lej	502	lek	503	lel	504	lem	505
len	506	leo	507	lep	508	leq	509	ler	510
les	511	let	512	leu	513	lev	514	lew	515
lex	516	ley	517	lez	518	lf	519	lg	520
lh	521	li	522	lj	523	lk	524	ll	525
lm	526	ln	527	lo	528	loa	529	lob	530
loc	531	lod	532	loe	533	lof	534	log	535
loh	536	loi	537	loj	538	lok	539	lol	540
lom	541	lon	542	loo	543	lop	544	loq	545
lor	546	los	547	lot	548	lou	549	lov	550
low	551	lox	552	loy	553	loz	554	lp	555
lq	556	lr	557	ls	558	lt	559	lu	560
lv	561	lw	562	lx	563	ly	564	lz	565
m	566	ma	567	maa	568	mab	569	mac	570
mad	571	mae	572	maf	573	mag	574	mah	575
mai	576	maj	577	mak	578	mal	579	mam	580
man	581	mao	582	map	583	maq	584	mar	585
margaret	586	mary	587	mas	588	mat	589		
mau	590	mav	591	maw	592	max	593	may	594
maz	595	mb	596	mc	597	md	598	me	599
mf	600	mg	601	mh	602	mi	603	mj	604
mk	605	ml	606	mm	607	mn	608	mo	609
mp	610	mq	611	mr	612	ms	613	mt	614
mu	615	mv	616	mw	617	mx	618	my	619
mz	620	n	621	na	622	nb	623	nc	624
nd	625	ne	626	nf	627	ng	628	nh	629
ni	630	nj	631	nk	632	nl	633	nm	634
nn	635	no	636	np	637	nq	638	nr	639
ns	640	nt	641	nu	642	nv	643	nw	644
nx	645	ny	646	nz	647	o	648	oa	649
ob	650	oc	651	od	652	oe	653	of	654
og	655	oh	656	oi	657	oj	658	ok	659
ol	660	om	661	on	662	oo	663	op	664

oq	665	or	666	os	667	ot	668	ou	669
ov	670	ow	671	ox	672	oy	673	oz	674
p	675	pa	676	pb	677	pc	678	pd	679
pe	680	pf	681	pg	682	ph	683	pi	684
pj	685	pk	686	pl	687	pm	688	pn	689
po	690	pp	691	pq	692	pr	693	ps	694
pt	695	pu	696	pv	697	pw	698	px	699
py	700	pz	701	q	702	qa	703	qb	704
qc	705	qd	706	qe	707	qf	708	qg	709
qh	710	qi	711	qj	712	qk	713	ql	714
qm	715	qn	716	qo	717	qp	718	qq	719
qr	720	qs	721	qt	722	qu	723	qv	724
qw	725	qx	726	qy	727	qz	728	r	729
ra	730	rb	731	rc	732	rd	733	re	734
rf	735	rg	736	rh	737	ri	738	rj	739
rk	740	rl	741	rm	742	rn	743	ro	744
robert	745	rp	746	rq	747	rr	748	rs	749
rt	750	ru	751	rv	752	rw	753	rx	754
ry	755	rz	756	s	757	sa	758	sb	759
sc	760	sd	761	se	762	sf	763	sg	764
sh	765	si	766	sj	767	sk	768	sl	769
sm	770	sn	771	so	772	sp	773	sq	774
sr	775	ss	776	st	777	su	778	sv	779
sw	780	sx	781	sy	782	sz	783	t	784
ta	785	tb	786	tc	787	td	788	te	789
tf	790	tg	791	th	792	ti	793	tj	794
tk	795	tl	796	tm	797	tn	798	to	799
tp	800	tq	801	tr	802	ts	803	tt	804
tu	805	tv	806	tw	807	tx	808	ty	809
tz	810	u	811	ua	812	ub	813	uc	814
ud	815	ue	816	uf	817	ug	818	uh	819
ui	820	uj	821	uk	822	ul	823	um	824
un	825	uo	826	up	827	uq	828	ur	829
us	830	ut	831	uu	832	uv	833	uw	834
ux	835	uy	836	uz	837	v	838	va	839
vb	840	vc	841	vd	842	ve	843	vf	844
vg	845	vh	846	vi	847	vj	848	vk	849
vl	850	vm	851	vn	852	vo	853	vp	854
vq	855	vr	856	vs	857	vt	858	vu	859
vv	860	vw	861	vx	862	vy	863	vz	864

w	865	wa	866	wb	867	wc	868	wd	869
we	870	wf	871	wg	872	wh	873	wi	874
wia	875	wib	876	wic	877	wid	878	wie	879
wif	880	wig	881	wih	882	wii	883	wij	884
wik	885	wil	886	william	887	wim	888		
win	889	wio	890	wip	891	wiq	892	wir	893
wis	894	wit	895	wiu	896	wiv	897	wiw	898
wix	899	wiy	900	wiz	901	wj	902	wk	903
wl	904	wm	905	wn	906	wo	907	wp	908
wq	909	wr	910	ws	911	wt	912	wu	913
wv	914	ww	915	wx	916	wy	917	wz	918
x	919	xa	920	xb	921	xc	922	xd	923
xe	924	xf	925	xg	926	xh	927	xi	928
xj	929	xk	930	xl	931	xm	932	xn	933
xo	934	xp	935	xq	936	xr	937	xs	938
xt	939	xu	940	xv	941	xw	942	xx	943
xy	944	xz	945	y	946	ya	947	yb	94
yc	949	yd	950	ye	951	yf	952	yg	953
yh	954	yi	955	yj	956	yk	957	yl	958
ym	959	yn	960	yo	961	yp	962	yq	963
yr	964	ys	965	yt	966	yu	967	yv	968
yw	969	yx	970	yy	971	yz	972	z	973
za	974	zb	975	zc	976	zd	977	ze	978
zf	979	zg	980	zh	981	zi	982	zj	983
zk	984	zl	985	zm	986	zn	987	zo	988
zp	989	zq	990	zr	991	zs	992	zt	993
zu	994	zv	995	zw	996	zx	997	zy	998
zz	999								

Middle Name. Use the table above to also code the middle name as 3 digits. If there is no middle name and the first name is fully coded (say, "John" or "Mary" or a full first name), then the digits for the middle name is 000. If the first name is not fully coded, encode the first unused character from the first name using the table below.

a	001	b	002	c	003	d	004	e	005
f	006	g	007	h	008	i	009	j	010
k	011	l	012	m	013	n	014	o	015
p	016	q	017	r	018	s	019	t	020
u	021	v	022	w	023	x	024	y	025
z	026								

Birth day and month. Look up the birth month and day on the following tables. In the event of two or more people have identical driver's license numbers, this final group of digits is also used to differentiate them. Simply add one to the final group of digits until there is an unused entry. If you reach a number allocated to a different date, subtract one until you find an unused entry.

January (001)
01 => 002 02 => 007 03 => 010 04 => 012
05 => 017 06 => 020 07 => 022 08 => 025
09 => 027 10 => 030 11 => 032 12 => 035
13 => 037 14 => 040 15 => 042 16 => 045
17 => 047 18 => 050 19 => 052 20 => 055
21 => 057 22 => 060 23 => 062 24 => 065
25 => 067 26 => 070 27 => 072 28 => 075
29 => 077 30 => 080 31 => 082

February (085)
01 => 086 02 => 088 03 => 091 04 => 093
05 => 096 06 => 098 07 => 101 08 => 103
09 => 106 10 => 108 11 => 111 12 => 113
13 => 116 14 => 118 15 => 121 16 => 123
17 => 126 18 => 128 19 => 131 20 => 133
21 => 136 22 => 138 23 => 141 24 => 143
25 => 146 26 => 148 27 => 151 28 => 153
29 => 156

March (158)
01 => 159 02 => 162 03 => 164 04 => 167
05 => 169 06 => 172 07 => 174 08 => 177
09 => 182 10 => 184 11 => 187 12 => 189
13 => 192 14 => 194 15 => 197 16 => 199
17 => 202 18 => 204 19 => 207 20 => 227
21 => 229 22 => 232 23 => 234 24 => 237
25 => 239 26 => 242 27 => 244 28 => 247
29 => 249 30 => 252 31 => 254

April (257)
01 => 258 02 => 260 03 => 263 04 => 265
05 => 268 06 => 270 07 => 273 08 => 275
09 => 278 10 => 280 11 => 283 12 => 285
13 => 288 14 => 290 15 => 293 16 => 295
17 => 298 18 => 300 19 => 303 20 => 305
21 => 308 22 => 310 23 => 313 24 => 315

25 => 318 26 => 320 27 => 323 28 => 325
29 => 328 30 => 330

May (333)
01 => 334 02 => 336 03 => 339 04 => 341
05 => 344 06 => 346 07 => 349 08 => 351
09 => 354 10 => 356 11 => 359 12 => 361
13 => 364 14 => 366 15 => 369 16 => 371
17 => 374 18 => 376 19 => 379 20 => 381
21 => 384 22 => 386 23 => 389 24 => 391
25 => 394 26 => 396 27 => 399 28 => 401
29 => 404 30 => 406 31 => 409

June (411)
01 => 412 02 => 415 03 => 417 04 => 420
05 => 422 06 => 425 07 => 427 08 => 430
09 => 432 10 => 435 11 => 437 12 => 440
13 => 442 14 => 445 15 => 447 16 => 450
17 => 452 18 => 467 19 => 470 20 => 472
21 => 475 22 => 477 23 => 480
24 => 482 25 => 497 26 => 500
27 => 502 28 => 505 29 => 507
30 => 517

July (520)
01 => 521 02 => 523 03 => 526 04 => 528
05 => 534 06 => 537 07 => 539 08 => 542
09 => 544 10 => 547 11 => 549 12 => 552
13 => 554 14 => 557 15 => 559 16 => 562
17 => 564 18 => 567 19 => 569 20 => 572
21 => 574 22 => 577 23 => 579 24 => 582
25 => 584 26 => 587 27 => 589 28 => 592
29 => 594 30 => 597 31 => 599

August (602)
01 => 603 02 => 605 03 => 608
04 => 610 05 => 613 06 => 615 07 => 618
08 => 620 09 => 623 10 => 625 11 => 628
12 => 630 13 => 633 14 => 635 15 => 638
16 => 640 17 => 643 18 => 645 19 => 648
20 => 650 21 => 653 22 => 655 23 => 658
24 => 660 25 => 663 26 => 665 27 => 668

28 => 670 29 => 673 30 => 675 31 => 678

September (680)
01 => 681 02 => 684 03 => 686 04 => 689
05 => 691 06 => 694 07 => 696 08 => 699
09 => 701 10 => 704 11 => 706 12 => 709
13 => 711 14 => 714 15 => 716 16 => 719
17 => 721 18 => 724 19 => 726 20 => 729
21 => 731 22 => 734 23 => 736 24 => 739
25 => 741 26 => 744 27 => 746 28 => 749
29 => 751 30 => 754

October (756)
01 => 757 02 => 760 03 => 762 04 => 765
05 => 767 06 => 770 07 => 772 08 => 775
09 => 777 10 => 780 11 => 782 12 => 785
13 => 787 14 => 790 15 => 792 16 => 797
17 => 800 18 => 802 19 => 807 20 => 810
21 => 812 22 => 815 23 => 817 24 => 820
25 => 822 26 => 825 27 => 827 28 => 830
29 => 832 30 => 835 31 => 837

November (840)
01 => 841 02 => 843 03 => 846 04 => 848
05 => 851 06 => 853 07 => 856 08 => 858
09 => 861 10 => 863 11 => 866 12 => 868
13 => 871 14 => 873 15 => 876 16 => 878
17 => 881 18 => 883 19 => 886 20 => 888
21 => 891 22 => 893 23 => 896 24 => 898
25 => 901 26 => 903 27 => 906 28 => 908
29 => 911 30 => 913

December (916)
01 => 917 02 => 919 03 => 922 04 => 924
05 => 927 06 => 929 07 => 932 08 => 934
09 => 937 10 => 939 11 => 942 12 => 944
13 => 947 14 => 949 15 => 952 16 => 954
17 => 957 18 => 959 19 => 962 20 => 964
21 => 967 22 => 969 23 => 972 24 => 974
25 => 977 26 => 983 27 => 985 28 => 990
29 => 993 30 => 995 31 => 998

A.3. Washington State Encoding of Driver License Number [133]

The checksum digit is computed after the rest of the number is known. a1 is the first character (the first letter of the last name), a2 is the second letter, and so on.

checksum = (a1 - a2 + a3 - a4 + a5 - a6 + a7 - a8 + a9 - a11 + a12) mod 10

For non digits, use the following values:

*	4	A	1	B	2	C	3	D	4	E	5
F	6	G	7	H	8	I	9	J	1	K	2
L	3	M	4	N	5	O	6	P	7	Q	8
R	9	S	2	T	3	U	4	V	5	W	6
X	7	Y	8	Z	9						

Look up your birth month in the following table. There are two options. In general, use the first column. If another person has generated the same license number, use the second column.

Jan	1	B	S
Feb	2	C	T
Mar	3	D	U
Apr	4	J	1
May	5	K	2
Jun	6	L	3
Jul	7	M	4
Aug	8	N	5
Sep	9	O	6
Oct	10	P	7
Nov	11	Q	8
Dec	12	R	9

Look up the code for the day of the birthday in the following table.

1	A	2	B	3	C	4	D	5	E	6	F
7	G	8	H	9	Z	10	S	11	J	12	K
13	L	14	M	15	N	16	W	17	P	18	Q
19	R	20	0	21	1	22	2	23	3	24	4
25	5	26	2	27	7	28	8	29	9	30	T
31	U										