

**A Technical Analysis of Issues Plaguing the Data and Information Technology Systems
of Political Campaigns in the United States**

A Technical Research Paper Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

By Liem Budzien, Kevin Dang

Fall 2020

On my honor as a University Student, I have neither given nor received
unauthorized aid on this assignment as defined by the Honor
Guidelines for Thesis-Related Assignments.

Signature *Liem Budzien* Date 11/30/2020
Liem Budzien

Approved *Nada Basit* (nb3f) Date 11/23/2020
Nada Basit, Department of Computer Science

Introduction

In the United States, there are roughly 235 million eligible voters. However, during most elections, despite the best efforts of campaigns to drive turnout, only about half of the eligible voters cast a ballot. In fact, in the November 2020 presidential election that had historic turnout, only ~68% of eligible voters cast a vote (Dottle and Pogkas, 2020). The level of voter turnout for a specific candidate or campaign can often tip the race in their favor—thus, vast sums of money are spent to identify and target voters. For the 2020 Presidential election cycle, an estimated \$11 billion has been spent between the campaigns and interest groups (Bruell, 2019). A large portion of this funding goes to infrastructure—the information and technology (IT) systems that function as the backbone for campaigns, from local to national in scope. They manage voter files, online ad campaigns, and volunteer canvassing. However, despite the deluge of funding, these technological systems historically struggle to support modern campaigns, often plagued by weak security, poor reliability, and high costs.

Between the two major political parties in the United States, it is widely accepted that the Republican data and IT operation is more sophisticated, with more accurate data and modern tools for digital campaigns and voter targeting. Meanwhile, the Democrat party has historically struggled over the past decade to modernize and build an operation capable of supporting various campaigns. Conversely, the Democratic party's online fundraising operation has been a runaway success, while the Republicans have continued to rely on individual, large-sum donors due to struggles fundraising online. Thus, the contrast between the two digital operations can serve as a case study to develop suggestions and recommendations regarding the best practices for building information technology infrastructure and data exchanges. Various elements of the operations are

examined, including funding sources, operating structures, and the specific technologies employed.

Risks of Weak Data Operations

Campaign technology has historically been behind the curve given the advancements in technology, especially information technology (IT). A contributing factor to the weakness of campaign IT systems is the practice of building a new system for every election cycle. This leads to the fragmentation of standards and prevents the continual development of robust infrastructure (Lapowsky, 2015). Without robust infrastructure, issues such as security and reliability arise. In 2017, the personal information of 198 million Americans was leaked from a campaign database, exposing information such as legal names, addresses, registered political parties, and racial demographics. In 2018, the voter registration information for 35 million voters was offered for sale on a dark web hacking forum (Wright, 2019). Meanwhile, a synopsis of Hillary Clinton's 2016 presidential bid found that a poor data operation and IT infrastructure hamstrung the effectiveness of the campaign. Research has shown that voter targeting technologies drive a statistically significant increase in voter turnout and influence voter decisions by identifying specific voters who are favorable to a candidate (Hoferer et al., 2019). However, robust IT infrastructure is required to power these data-intensive technologies.

Thus, this paper seeks to answer the question: *How have political entities failed to create robust, secure data and information technology systems, and what steps should be taken to modernize the way campaigns build information technology infrastructure?* This research project's goal is to offer suggestions for creating a more secure, reliable, and transparent ecosystem of information technology infrastructure for campaigns.

Background

Historical Struggles of the Democratic Party

Democratic campaigns, for the most part, have relied on pencil-and-paper to conduct voter targeting and turnout operations (Glazer, 2020). Due to a lack of infrastructure to process and manage canvassing digitally, most volunteers are handed clipboards with a printed list of names and addresses. Furthermore, much of the data is of poor quality—as Democrats lack a central data exchange platform, updated data is often siloed into individual repositories which most campaigns lack access to. In contrast, the Republican data operation is much more sophisticated and better funded. Volunteers often go door-to-door wielding an iPad running the i360 app, a product developed to handle all aspects of canvassing (Ryan-Mosley, 2020). The app informs canvassers which doors to knock on, personalizes the questions based on each household, and immediately uploads the results of the survey to a centralized database in the cloud. The lack of an equivalent platform for Democrats have hampered efforts to increase voter turnout, most recently witnessed by the COVID-19 pandemic. Without an easy or safe way to reach voters in person, online methods of voter outreach have increased in importance. As the 2020 campaign season has come to a close, it is more apparent than ever that political parties must hold the ability to drive turnout using alternative methods to in-person canvassing.

Despite focusing earlier on social media and online targeting of voters in the 2004 and 2008 elections, the Democratic party has fallen behind on data information and collection regarding individual voter information (Kreiss, 2019). Following the 2008 and 2012 victories of Barack Obama, the Democratic candidate, the party failed to modernize the backend infrastructure used for data science and online engagement (Glazer, 2020). The original technology stack was named Vertica, a data repository system maintained by the Democratic

National Committee (DNC) which was used by many Democratic organizations (Epstein, 2020). However, the original system was fragmented—each campaign and organization contained data and voter information within disparate databases, without sharing or storage within a central system (Epstein, 2020). Thus, despite a shared technology backend platform, none of the data was pooled and any benefits of centralization for data analysis were lost. Furthermore, built on outdated technology, as the complexity of requirements grew, performance and reliability suffered. As Michael Slaby, the chief innovation officer for Obama’s 2012 campaign stated, “What [the Clinton] campaign was reacting to was a failure over the course of the Obama years to effectively keep up the pace of modernization inside the DNC. Technology doesn’t sit still for 10 years” (Lapowsky, 2019). Vertica was famous for its shortcomings, such as disorganized and unlabeled data. In fact, the 2016 campaign team for Hillary Clinton employed a 24-hour technology team to simply work constantly to keep the system functioning. If too many requests were sent to the servers, the system was known to crash and be down for 16 hours at a time—valuable time during an election and the final weeks of a campaign (Lapowsky, 2019).

The starkest example was 2016, where Clinton’s loss was widely blamed on a lack of online advertising, particularly Facebook. In addition to the Koch Network’s data science team, the Trump campaign employed organizations such as Cambridge Analytica to collect data and micro-target potential voters online (Ratnam, 2020). The data included the cell-phone numbers harvested from tracking the participants in right-wing rallies and what Facebook pages users liked in order to build a psychological profile of voters online. These detailed profiles, compiled from data non-traditionally associated with voter profiles, enabled the Trump campaign and the Republican party to serve micro-targeted advertisements designed to appeal to each voter’s specific interests (Sullivan, 2020). Instead of running traditional television advertising using a

selected number of high-production ads, the Trump campaign generated 5.8 million unique ads that were run online, on Facebook in particular. Compared to the 66,000 unique advertisements produced by the Clinton campaign, the far greater number of unique advertisements allowed the campaign to target more specific voters, embracing the “direct to consumer” ideology of politics (Sullivan, 2020). This ability led to higher online engagement and more successfully converted potential votes into ballots cast on election day. Despite a loss in the popular vote, the Trump campaign managed to drive an unexpectedly high level of turnout for specific demographic groups, particularly non-college educated white males in the Midwest, handing Trump a victory in the Electoral College (Meko et. al., 2016).

Modern Day Data Operations

Over the last few years, campaigns have been steadily adding to the vast amount of personal information they keep on voters. That’s partly a result of a practice called acquisition advertising, in which campaigns run direct response ads that seek to get either contact information or opinions straight from an individual. For example, some political advertisements on Facebook will direct you to respond to a poll or fill out a form based on a topic. As of May, both presidential campaigns were spending upwards of 80% of their ad budgets on direct response ads (Ryan-Mosley, 2020). Another form of advertising popular among campaigns is microtargeting. An established form of advertising most recognize as Facebook ads, microtargeting enables campaigns to target smaller groups of voters on social media platforms based on their user profiles (Ryan-Mosley, 2020). Ad campaigns can be targeted toward specific groups on most online platforms such as Facebook, Twitter, or YouTube. Audience for these campaigns can be fine tuned by targeting a specific age group, ethnicity, and/or zip code

(Merrill, 2016). Outside of more general demographics, most platforms allow more granular targeting, for example, job profession, hobbies or interests, and previously liked posts. However, in order to know what specific audience to target, campaigns must analyze the data on voters. The main system that allows for the success of these ad-based campaigns are data exchanges.

The Republican Party has had data exchanges set up since 2013 starting with Data Trust, a central data exchange portal which has led to the establishment of the Republican Party as the technology powerhouse during campaigns. A data exchange allows companies, Political Action Committees (PACs), and campaigns to exchange and sync voter information between each other, with some voters having up to 2,500 data points in their profile (Ryan-Mosley, 2020). For example, during the 2018 midterm elections, Republicans used Data Trust to identify early voters and in turn stopped targeted advertising toward that population of voters, saving themselves \$100 million in expenses that would have otherwise been wasted without the usage of a data exchange and voter profiles (Ryan-Mosley, 2020). On the other hand, the Democrats have failed to establish their own form of a data exchange, which has led some to discern as the reason Hillary Clinton lost in 2016, as often explained by Clinton herself (Ryan-Mosley, 2020). The Democratic Party and its technology had stagnated under the second term of President Obama, leading to the large discrepancy between the technological capabilities between the two campaigns as Democrats were left competing against a much more modern Republican data operation. Unable to take advantage of recent advancements in data science and machine learning, the fragmented and fragile state of the Vertica system used by Democrats impeded efforts to build similarly complex and accurate voter profiles. Many of the benefits that a party-wide data exchange would have brought were lost out on until recently; the Democrats have just launched their version of Data Trust named Phoenix in September 2020, seven years

after the Republican Party first established theirs (Epstein, 2020). The seven year technology gap between the two parties has often made the Democratic Party the lesser of the two in the ongoing arms race of voter data technology; the discrepancy resulting in major losses as demonstrated in 2016.

Implications of Weak Cybersecurity

As campaigns both hold the personal information of millions of Americans and are political bodies that influence the government, they are prime targets for cybersecurity threats. Political campaigns are effectively start-ups in terms of operations, especially in technology—the issue lies in the fact that campaigns are often built-up from the ground for every new election cycle (Bond, 2020). New campaigns involve a lot of traffic in-and-out of the employee and technical infrastructure: new employees bring new technologies with them creating risks in security for both political parties. Furthermore, with a high volume of communications from both professional and personal devices, there are many opportunities for cracks in the security to become vulnerable to cyber attacks. One of the widest known examples being of Hillary Clinton’s campaign chair, John Podesta, having his personal Gmail account compromised by a phishing attack exposing private communications that were damaging to the campaign and its image (Bond, 2020). Meanwhile, during the 2020 election cycle, the Wisconsin Republican Party lost \$2.3 million to hackers who gained access to the organization’s financial accounts.

With such large risks at play in political campaigns, it further supports the idea of developing secure infrastructure for data and communications. Steps to counteract these threats have been taken in campaigns since these attacks, including stricter protocols for online communications. For example, Google provides physical security keys used for Two Factor

Authentication (2FA), which can be used to secure login credentials for email or other online accounts (Newman, 2020). In addition to providing additional security measures, the organizational structure of the IT operation also plays an important role. Without dozens of separate databases and systems, a single centralized repository for data and analysis, such as Phoenix and Data Trust, minimize the number of opportunities for cybersecurity threats to gain access. Storing the databases and performing any computation or analysis in the cloud also improves security, with Phoenix relying on the Google cloud platform for data storage and Google's BigQuery for data analysis (Lapowsky, 2019). By utilizing cloud services from established providers, security is improved over attempting to secure a custom implementation. Google and other cloud providers are able to invest much more time and resources into securing their platforms, removing a portion of the burden from individual campaigns.

Democrats seek to modernize their infrastructure

Following the unexpected loss in 2016, Democrats quickly mobilized to modernize their outdated campaign technology infrastructure, particularly regarding data collection and voter targeting. In recent years, a surge in support for the Democratic party from technology hubs in the United States, particularly Silicon Valley, has enabled the party to make strides in assembling a more modern infrastructure backbone. Alloy, a Democratic party aligned data science startup, was founded by Reidd Hoffman, the billionaire LinkedIn founder as a response to the Koch brothers' conservative action group (Sullivan, 2020). Alloy developed a cloud-based data sharing platform from the ground up with reliability and security in mind, opening up a data exchange for Democratic groups to contribute to a centralized pool of voter information (Sullivan, 2020). In addition, Alloy has funded the creation of apps intended to boost voter turnout, such as Hustle

and Peer, which provide voting information and peer-to-peer messaging services for political campaigns (Sullivan, 2020).

Democratic successes and challenges for the Republican party

In order to fund significant voter turnout operations and advertising, especially with the notoriously expensive media markets within the United States, the parties rely on raising money from political donors. Traditionally, a small number of wealthy individuals would donate large dollar amounts to fund campaigns. However, modern campaigns have begun to increasingly rely on small-dollar donations from many individual donors, particularly through online donations. One area where Democrats have outpaced their Republican counterparts has been online fundraising. A central platform, ActBlue, processes transactions for any Democratic campaign and organization. Recent examples of the platform's success have been the death of Ruth Bader Ginsburg and the announcement of Kamala Harris as Vice President, where liberal groups raised tens of millions of dollars in a single night (Goldmacher, 2020). Over the third calendar year quarter, ActBlue announced they processed over \$1.5 billion in online donations, mainly from small-dollar individual contributions (Goldmacher, 2020). In contrast, the Republican party developed a centralized online fundraising platform years after ActBlue. Their equivalent organization, WinRed, has so far refused to release numbers that would lead to comparisons. Democrats have taken advantage of their vast email lists and individual online donors to dramatically out raise the Republican party for nationwide and local campaigns (Gratzinger, 2020). This fundraising advantage has put Republicans on the defensive for the upcoming elections.

Looking Toward the Future

For the future of political campaigns, utilizing the advantages of Big Data and data analytical tools requires an enormous amount of information to produce more accurate results. Regarding current usage of Big Data, both parties have failed to modernize their technology to a point where it can match other data powerhouses. The Democratic Party's old system named Vertica was one of many failures that had little promise if compared to technology today. Vertica's uptime would be around 50% as its servers would go down every other day due to the large influx of users and their information (Epstein, 2020). Vertica held information across many different databases which hampered performance and increased security vulnerabilities as information networks are only as strong as the weakest (Epstein, 2020). If one database were to be hacked, it would become a liability issue for the entire system. As of right now, the DNC has upgraded their infrastructure to a new data warehouse known as Phoenix. Phoenix is hosted by Google Cloud and supported using Google Analytics, a tool many recognize as the forefront in data engineering and analytical technology (DNC, 2020). This new system is utilized by state parties, sister committees, and candidate campaigns. Phoenix offers new benefits including "99.9% uptime, essential security controls, and near infinite scalability" (DNC, 2020). With the modernization of Democrat IT, the Republicans now find themselves behind in centralization. The Democrats have always had ActBlue which served as a website that centralizes all acts of funding towards the DNC. The streamlined process in donating comes with the benefit of building familiarity and establishing habits with donors which enables the Democratic Party to raise more money online.

Conclusion

While neither party holds a clear advantage in technological infrastructure, with Republicans wielding a more advanced data operation and Democrats holding a vastly more successful online fundraising platform, the common thread is clear. Centralization and simplification are the key to robust, secure data and information technology systems. Through building a single system and shared infrastructure for donation and voter data, the political parties in the United States have established advantages in each area. Thus, as the Democrats launch Phoenix and the Republicans launch WinRed, they are seeking to catch up with their peers. Utilizing technologies such as BigQuery from Google for data analytics, the more data stored in a centralized database, higher quality and deeper insights can be developed. As elections shift into the digital age, micro-targeting of voters is paramount. As evidenced by the deluge of technology startups focused on peer-to-peer texting and apps for voter canvassing, the technology stack held by campaign organizations contributes to a distinct advantage in voter turnout. As the parties have worked to remedy historical flaws in their operations, particularly outdated and fragmented backend systems, the ability to centralize and simplify infrastructure improves security, reliability, and efficiency. In the United States, as election day nears, the ability to mobilize the 235 million eligible voters is critical to preserving democracy and encouraging citizens to make their voices heard.

References

- Baldwin-Philippi, J. (2017). The Myths of Data-Driven Campaigning. *Political Communication*, 34(4), 627–633. <https://doi.org/10.1080/10584609.2017.1372999>
- Bond, S. (2020, January 28). *2020 Political Campaigns Are Trying To Avoid A 2016-Style Hack*. NPR.
<https://www.npr.org/2020/01/28/799062773/2020-political-campaigns-are-trying-to-avoid-a-2016-style-hack>.
- Dottle, R., & Pogkas, D. (2020, November 4). *Voter Turnout Hits Historic Levels With States Still Counting Votes*. Bloomberg.
<https://www.bloomberg.com/graphics/2020-us-election-voter-turnout/>.
- Epstein, R. J. (2020, September 6). *Democrats Belatedly Launch Operation to Share Information on Voters*. The New York Times.
<https://www.nytimes.com/2020/09/06/us/politics/Presidential-election-voting-Democrats.html>.
- Glazer, E., & Seetharaman, D. (2020, August 28). *Democrats Lag Behind GOP in Data Race, Despite Making Progress*. The Wall Street Journal.
<https://www.wsj.com/articles/democrats-lag-behind-gop-in-data-race-despite-making-progress-11598615530>.
- Goldmacher, S. (2020, September 19). *Democrats shatter ActBlue's donation records after Ginsburg's death*. The New York Times.

<https://www.nytimes.com/2020/09/18/us/elections/democrats-shatter-actblues-donation-records-after-ginsburgs-death.html>.

Gratzinger, O. (2020, October 30). *Small donors give big money in 2020 election cycle*.

OpenSecrets.org.

<https://www.opensecrets.org/news/2020/10/small-donors-give-big-2020-thanks-to-technology/>.

Hern, A. (2018, May 6). *Cambridge Analytica: how did it turn clicks into votes?* The Guardian.

<https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie>.

Kreiss, D. (2019, April 30). *The Democratic advantage in digital, data, and analytics isn't going away easily*. The Washington Post.

<https://www.washingtonpost.com/news/monkey-cage/wp/2015/05/13/the-democratic-advantage-in-digital-data-and-analytics-isnt-going-away-easily/>.

Kreiss, D., & Jasinski, C. (2016). The Tech Industry Meets Presidential Politics: Explaining the Democratic Party's Technological Advantage in Electoral Campaigning, 2004–2012. *Political Communication*, 33(4), 544–562.

<https://doi.org/10.1080/10584609.2015.1121941>

Lapowsky, I. (2019, April 2). *Inside the Democrats' Plan to Fix Their Crumbling Data Operation*. Wired.

<https://www.wired.com/story/democrats-fix-crumbling-data-operation/>.

- Lapowsky, I. (2015, October 28). *Political Campaigns Are Wasteful-So Turn Them Into Startups*. Wired. <https://www.wired.com/2015/10/2016-election-tech/>.
- Meko, T., Lu, D., & Gamio, L. (2016, November 11). *How Trump won the presidency with razor-thin margins in swing states*. The Washington Post. <https://www.washingtonpost.com/graphics/politics/2016-election/swing-state-margins/>.
- Merrill, J. B. (2016, November 7). *Why Facebook Showed You That Ad for the Candidate You Hate*. The New York Times. <https://www.nytimes.com/2016/11/08/us/politics/facebook-ads-campaign.html>.
- Newman, L. H. (2020, February 11). *Google's Giving Out Security Keys to Help Protect Campaigns*. Wired. <https://www.wired.com/story/google-free-security-keys-campaigns/>.
- Ratnam, G. (2020, July 29). *Republicans and Democrats battle over who has better voter data effort*. <https://www.rollcall.com/2020/07/29/republicans-and-democrats-battle-over-who-has-better-voter-data-effort/>.
- Ryan-Mosley, T. (2020, October 3). *The technology that powers the 2020 campaigns, explained*. MIT Technology Review. <https://www.technologyreview.com/2020/09/28/1008994/the-technology-that-powers-political-campaigns-in-2020-explained/>.
- Sullivan, M. (2020, February 3). *Trump's data team crushed Democrats in 2016. Can this small startup even the score?*Fast Company.

<https://www.fastcompany.com/90454326/trumps-data-team-crushed-democrats-in-2016-can-this-small-startup-even-the-score>.