# Taking the Law into our Own Hands: Effective Data Security through User Education

Christopher Lindsey

School of Information, San Jose State University

INFM 203: Big Data Analytics and Management

Dr. Glen Mules

December 10, 2020

## Abstract

As the number of yearly data security breaches continue to rise, it is becoming more and more important for businesses, engineers, and individuals to embrace effective security practices. Traditional security approaches focus on what could be called a “computer scientist vs. evil computer scientist”-model that may be less relevant now as recent bad actors cast a wide net of phishing emails and brute-force attacks with the intention of getting users to mistakenly give up their authentication credentials so that these bad actors may simply log in to an otherwise very secure network. The solution to this problem does not lie in more robust firewalls and more advanced threat detection systems but in simple user education and continued, basic cybersecurity testing for ALL employees and network users.

Part II covers ethical questions in cybersecurity. Specifically, what individuals and businesses must do when the laws have not been written and the tools have not been developed to protect us from the unscrupulous actors among us. The solution, again, lies in user education.

# Taking the Law into our Own Hands: Effective Data Security through User Education

## Part I: Building the Cybersecurity Solution

According to the Identify Theft Resource Center, data breaches rose by over 100% between 2010 and 2019, from 662 in 2010 to 1,473 in 2019 (as cited by iii.org, 2019). The number of records exposed during that time period rose over ten-fold, from 16.2 million in 2010 to 164.7 million in 2019 (as cited by iii.org, 2019). In short, effective data governance is as relevant a topic as ever and getting more relevant each year. In this essay, I will discuss a subset of data governance—data security governance. So, what is data security governance?

Though the original is trapped behind a paywall, Gartner Research is often cited defining data security governance as "a subset of information governance that deals specifically with protecting corporate data (in both structured database and unstructured file-based forms) through defined data policies and processes" (as cited in Davis, 2020).

*Author's note*: Though there are numerous subtle yet important distinctions between them, in the context of this essay, I will use the terms "data" and "information" mostly interchangeably.

### The Five Pillars of Data Security

In Cloudera's documentation on Hadoop security for the Hortonworks Data Platform, they provide a theoretical framework for "a holistic approach" to data security "that revolves around five pillars of security: administration, authentication and perimeter security, authorization, auditing, and data protection" (HDP, n.d.). In this essay, I will discuss the first two: administration and authentication. Or, another way to describe this topic is: setting policies to enable more effective authentication.

**CIA Triad**

No essay about data security governance would be complete without a nod to the CIA triad. The CIA Triad provides three essential tenets of information security: confidentiality, integrity, and availability. In this essay, I will primarily focus on the first: confidentiality.

In "The CIA Strikes Back: Redefining Confidentiality, Integrity and Availability in Security," Coss and Samonas review the history of the CIA triad, noting its origins in "The Anderson Report," which was commissioned by the US Air Force in 1972. "The Anderson Report" identified "three categories of potential security risks that eventually became the foundation of the CIA triad" (Coss, Samonas, 2014):

1) Unauthorized information release: an unauthorized person is able to read and take advantage of information stored in the computer. This category of concern sometimes extends to "traffic analysis," in which the intruder only observes the patterns of information use. From those patterns, the intruder can infer some information content. This category also includes the unauthorized use of a proprietary program. (Confidentiality)

2) Unauthorized information modification: an unauthorized person is able to make changes in stored information – a form of sabotage. It should be noted that in the case of this kind of violation, the intruder does not necessarily see the information he has changed. (Integrity)

3) Unauthorized denial of use: an intruder can prevent an authorized user from referring to, or from modifying information, even though the intruder may not be able to refer to, neither modify the information themselves. (Availability)

(as cited in Coss, Samonas, 2014)

Again, this essay focuses on ways to maintain confidentiality and prevent, as "The Anderson Report" calls it, "unauthorized information release" (as cited in Coss, Samonas, 2014).

## Traditional Methods of Network Security

In *Hadoop Security* by Ben Spivey and Joey Echeveria, Spivey and Echeveria recommend two approaches to network security: segmentation and intrusion detection & prevention.

### *Segmentation*

Spivey and Echeveria (2015) discuss two types of segmentation: physical and logical network segmentation. Physical segmentation "is achieved by sectioning off a portion of the network with devices such as routers, switches, and firewalls" (p. 32). Logical segmentation is achieved through the use of unique Internet Protocol (IP) addressing (p. 32).

Spivey and Echeveria (2015) also recommend implementing network firewalls with carefully-granted permissions: "Network firewalls are a great way to enforce separation of a Hadoop cluster from the rest of the network that it resides in" (p. 33). They denote three types of firewall protections: one for data movement, another for client access, and another for administration traffic (Spivey, Echeveria, 2015, p. 39). They encourage network administrators to keep master, worker, and management nodes limited to administrators (pp. 39–42).

Spivey and Echeveria encourage network administrators to keep a close eye on their users and only allow them direct access to what they need. By preventing users from directly accessing high-risk master and worker nodes in a cluster, you prevent users from intentionally or accidentally engaging in activities which may cause a DDoS or expose data which ought not be exposed.

***Intrusion detection & prevention***

"Now that we have a Hadoop cluster that resides on its own network segment, how is this segment protected? This is largely achieved with network firewalls, and intrusion detection and prevention systems" (Spivey, Echeveria, 2015, p. 33). Spivey and Echeveria discuss two types of protection systems: Intrusion detection systems (IDS) and instrusion prevention systems (IPS).

Intrusion detection systems (IDS) are designed to notify network administrators of unusual users or behaviors on a network. Spivey and Echeveria (2015) note that "placing an IDS inside the trusted network can be a valuable tool to warn administrators against the insider threat" (p. 37). Intrusion prevention systems (IPS) take it a step further by preventing intrusions. Spivey and Echeveria (2015) say that "placing an IPS in the ingest path of an Internet data source is perfectly reasonable" (p. 37).

**DBIR 2020: How is data security currently being breached?**

Germaine to the question of how to prevent data security breaches is the question of how is data security currently being breached? To get current answers to this question, one must look no further than Verizon's *Data Breach Investigations Report* (DBIR) (2020).

According to the DBIR (2020), the most common by far (45%) security breach tactic used in 2019 was hacking (p. 7). The DBIR (2020) defines hacking as "falling into three distinct groups: 1) those utilizing stolen or brute-forced credentials; 2) those exploiting vulnerabilities; and 3) attacks using backdoors and Command and Control (C2) functionality" (p. 19). And, shockingly, "Over 80% of breaches within Hacking involve Brute force or the use of lost or stolen credentials" (p. 19).

For reference, a Brute-force attack is defined on Wikipedia thusly: “In cryptography, a brute-force attack consists of an attacker submitting many passwords or passphrases with the hope of eventually guessing a combination correctly” (Brute-force attack, 2020).

**The Most Important Security Measure: Educating Users**

Spivey and Echeveria operate under the assumption that the leading cause of data security breaches are other computer scientists just like them, perhaps sitting right next to them. They state: “The Internet might be a hotbed of malicious actors, but the insider threat to a business is also very real and should not be overlooked. Security architecture choices must always be made under the assumption that the malicious actor works inside the same building you do” (Spivey, Echeveria, 2015, p. 37). However, perhaps because times have changed since 2015 when *Hadoop Security* was published (see “Mules’ Law”), the authors of the 2020 DBIR state to the contrary, that “it is a widely held opinion that insiders are the biggest threat to an organization’s security, but one that we believe to be erroneous.” (p. 10). In an analysis of the 2020 DBIR, Thycotic noted that only “8% of breaches were misuse by authorized users” (Carson, 2020). Thycotic’s Joseph Carson goes on:

> Email continues to be the top delivery method and office attachments, again the top payload with Web Applications, Desktops/Laptops and Email being the top target assets…Cyber criminals target your personal data and credentials. Your email is essentially your digital identity, and once a criminal has access to your email they can steal your identity and become you” (2020)

It is clear from the research in the DBIR that the greatest current threats to network and data security are things like easy-to-guess passwords and phishing emails. The only solution to these lies in user and employee education (and multifactor authorization, too).

There are many articles and approaches to educating users and employees on being more cyber-secure. Rather than argue for one pedagogical approach over another, I'll simply point out a key takeaway from Stephanie Baskerville's article, "Cybersecurity 101: Your end-users are the first line of defense" (Baskerville, 2018).

Baskerville recommends building the cybersecurity strategy around educating end-users (2018). She suggests using "a tool that creates a fake phishing email and see how many of your end-users open it" (Baskerville, 2018). Having worked there, I can attest that Facebook Inc uses this approach for continuous testing of their employees. Baskerville also encourages continuing education and refresher courses to keep employees ever vigilant.

Once your education system has been put in place, the U.S. Federal Trade Commission has several great "Cybersecurity Quizzes" on topics such as ransomware and phishing which can be used to continually test employees and network users to keep data safe and secure: https://www.ftc.gov/tips-advice/business-center/small-businesses/cybersecurity/quiz

## Part II: Ethics Surrounding Data Security

According to the BBC's "Ethics: a general introduction," "Ethics is concerned with what is good for individuals and society" (BBC, 2014). But what about when society doesn't know enough—about, say, a new form of technology—to make a sound distinction about what is good or bad for individuals and society. This ethical uncertainty has been a major aspect of Facebook's 2010s Cambridge Analytica data scandal.

According to Wikipedia's "Facebook–Cambridge Analytica data scandal" page, an investigation by the UK Information Commissioner's Office concluded that "the investigation did not identify any data misuse by Cambridge Analytica and concluded that their methods employed were common and well-recognized and were based on widely available technology" (Facebook–Cambridge Analytica data scandal, 2020). This data was acquired because around 300,000 people willingly downloaded an app called "This is your digital life" which provided the app owner—who turned out to be Cambridge Analytica—with a bunch of their Facebook data (W.I.R.E.D. Staff, 2018) as well as—thanks to Facebook's Open Graph platform—a whole bunch of all their friends' Facebook data. And all of this was done with the intention of targeting political ads.

What follows are many questions with grey-area answers. For example, while we can all agree that exposing someone's private data is unethical, do we also agree that targeting ads is unethical? In the case of Cambridge Analytica, their political ad targeting was the result of an arguable misuse of private and personal data which people uploaded onto Facebook without the knowledge that it might one day be used to sway an extremely contentious, divisive, and controversial U.S. election. So, to ask the question again in a slightly different way, is it unethical for a company to ask Facebook or Google to show that company's ads to those demographics which are most likely to buy that company's product? At first glance, the answer is a "probably not," especially since this is the source of most of Facebook and Google's revenue. But what about when one considers that Facebook and Google's own ad targeting is the result of privacy breaches which, frankly, most people are not fully aware of? Most people do not read the fine

print and do not know that Facebook "owns" all of your content, as soon as you click "post" or "upload."

These grey-area questions are in the process of being answered right now by society at large. With the technological landscape changing so quickly, ethical implications of a CEO's or an engineer's decisions do not become clear until the results of those decisions have fully played out. While many people might not see targeted ads as an ethical concern, once the targeting of those ads leads to an undesirable outcome, they might see the situation for what it is and see the actions that led to that undesirable outcome as having been unethical. Maybe the people who uploaded personal information to Facebook were not aware that one day one of their friends would download Cambridge Analytica's now-infamous "This is your digital life" app. The solution, once again, lies in education and in the general dissemination of ethical practices related to information storage and retrieval to compensate for instances where the law has failed us or where the law has not yet been written. As Dr. Hamid Nemati says, "It is safe to conclude that laws will not protect us, but can we seek the salvation in technology? The answer is no. Technology by itself is not the solution. Technology should be viewed as an enabler. It is the people who use the technology that are responsible to its ethical use" (Nemati, 2007, p. lxvi).

In "Information Security and Ethics: Concepts, Methodologies, Tools, and Applications" by Dr. Hamid Nemati, Nemati (2007) discusses the cost of privacy and why it matters. "Privacy concerns are real and have profound and undeniable implications on people's attitude and behavior" (p. lxx). Nemati (2007) notes that when consumers perceive that their privacy has been violated—like many Facebook users felt after the Cambridge Analytica scandal came to light—then they are "likely to lose confidence in the online marketplace because of these privacy

concerns" (p. lxx). And if people feel less trust in online marketplaces, then they're less likely to use them, and the e-commerce industry has to start laying people off, markets drop, the rich get less richer, and certainly no one wants that.

In addition to fiscal concerns, distrust erodes away the very fabric of an organized society. "Privacy is no longer about just staying within the letter of the latest law or regulation. As sweeping changes in attitudes of people their privacy will fuel an intense political debate and put once-routine business and corporate practices under the microscope" (Nemati, 2007, p. lxx). One can certainly feel the effects of the erosion of trust in business and government institutions in the debates over the outcomes of the 2020 U.S. election. Thus, in light of changing and emerging understanding of technology's impact on society, we must conclude, as Nemati does, that "organizations need to focus on information security not just as an IT issue, but rather as a business imperative" (Nemati, 2007, p. lxx).

**Conclusion**

The future is in our hands. We the people must educate ourselves about what it really means to upload data to Facebook or to some random app. And our great tech companies will be behooved more by a carefully and thoughtfully educated workforce than by anything else.

## References

*2020 Data Breach Investigations Report*. Verizon Enterprise. (2020). https://enterprise.verizon.com/resources/reports/dbir/.

*Archived Graphs: Number Of Data Breaches And Records Exposed, 2010-2019*. iii. (2020). https://www.iii.org/graph-archive/213434.

Baskerville, S. (2019, March 22). *Cybersecurity 101: Your End-Users are the First Line of Defense*. ProServeIT. https://www.proserveit.com/blog/end-user-education-best-cybersecurity-defense.

Carson, J. (2020, June 19). *Verizon DBIR: Top 5 Takeaways from the 2020 Report*. Thycotic. https://thycotic.com/company/blog/2020/06/17/verizon-2020-dbir-5-top-takeaways.

*Cybersecurity Quizzes*. Federal Trade Commission. (2020, March 23). https://www.ftc.gov/tips-advice/business-center/small-businesses/cybersecurity/quiz.

Davis, D. (2020, January 8). *Data Security Governance*. IRI. https://www.iri.com/blog/vldb-operations/data-security-governance/.

*HDP Documentation*. HDP. https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/bk_security/bk_security.pdf .

Nemati, Hamid. (2007). Information Security and Ethics: Concepts, Methodologies, Tools, and

Applications.

Samonas, S., & Coss, D. (1970, January 1). *The CIA Strikes Back: Redefining Confidentiality, Integrity and Availability in Security*. Journal of Information System Security. http://www.jissec.org/Contents/V10/N3/V10N3-Samonas.html.

Spivey, B., & Echeverria, J. (2015). *Hadoop security*. O'Reilly.

Staff, W. I. R. E. D. (2018, March 22). *The Cambridge Analytica Story, Explained*. https://www.wired.com/amp-stories/cambridge-analytica-explainer/.

Wikimedia Foundation. (2020, December 9). *Facebook–Cambridge Analytica data scandal*. Wikipedia. https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal.