

Hochschule für Technik und Wirtschaft des Saarlandes University of Applied Sciences



# Major Data Leaks from 2010–2019: The Snowden and WikiLeaks Era

Joel Alexandre Vogt
Technical Report – STL-TR-2020-03 – ISSN 2364-7167





Technische Berichte des Systemtechniklabors (STL) der htw saar Technical Reports of the System Technology Lab (STL) at htw saar ISSN 2364-7167

Joel Alexandre Vogt: Major Data Leaks from 2010-2019: The Snowden and WikiLeaks Era

Technical report id: STL-TR-2020-03

First published: July 2020 Last revision: September 2020 Internal review: André Miede

For the most recent version of this report see: https://stl.htwsaar.de/

Title image source: Republica, https://pixabay.com/photos/camera-graffiti-security-cctv-89012/



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. http://creativecommons.org/licenses/by-nc-nd/4.0/

# Major Data Leaks from 2010 - 2019: The Snowden and WikiLeaks Era

Joel Alexandre Vogt htw saar – Hochschule für Technik und Wirtschaft des Saarlandes Seminar "Computer Science and Society" Sommersemester 2020

Abstract—In the digital age, data leaks arise from seemingly nowhere and reveal major secrets – from a nation spying on its own citizens to an array of public figures committing fraud – on a large scale in a matter of minutes. Since data leaks typically consist of thousands of documents, journalists need months to investigate and publish an analysis of the events. This paper addresses the main data leaks from 2010-2019 – specifically, the NSA leak from whistle-blower Edward Joseph Snowden, the tax evasion leak outlined in The Panama Papers and CIA's hack toolbox, Vault 7.

#### I. INTRODUCTION

Over the last decade, there have been many data leaks on covert subjects ranging from surveillance programs such as the National Security Agency's (NSA) Planning Tool for Resource Integration, Synchronization and Management (PRISM) program – to offshore entities such as The Panama Papers. The act of leaking digital data is defined as 'allowing secret information to become generally known' [1] or 'to disclose without authorization or official sanction' [2]. In other words, a data leak is an intentional disclosure of secret information to the general public. In April 2010, the first major release from WikiLeaks captured the attention of the mainstream media with a video titled 'Collateral Murder' a classified US military video in which Iraqi citizens and Reuters news staff were killed by the US Army [3]. In 2013, whistle-blower Edward Joseph Snowden released thousands of classified documents through the British newspaper The Guardian and the American newspaper The Washington Post, which revealed the mass global surveillance being performed by the NSA. Three years later, another 11.5 million classified documents were leaked by the Süddeutsche Zeitung through an anonymous source about the offshore entities [4]. In 2017, WikiLeaks started the series of leaks known as 'Vault 7', which included documents discussing the activities and capabilities of the US Central Intelligence Agency (CIA) with regards to compromising electronic devices with malware [5]. This article will provide an overview and discussion of some of the major data and document leaks that occurred between 2010 and 2019. First, we will discuss the life of Edward J. Snowden, his activity as a whistle-blower and the aftermath of his actions. Second, we will examine the 2016 leak of 'The Panama Papers' - the largest data-leak to date. Finally, we will discuss Vault 7, which was released by WikiLeaks as part of a series of CIA leaks in March 2017. Due to the large volume

of leaked classified documents, this paper will investigate only the most significant leaks of the last decade.

# II. THE NSA & EDWARD JOSEPH SNOWDEN, WHISTLE-BLOWER

Edward J. Snowden is a computer scientist who became the most wanted man in the world following his release of NSA files to The Guardian and The Washington Post. James Clapper, the former director of the NSA, stated that Snowden 'probably downloaded' up to 1.7 million documents and revealed approximately 200,000 of these documents to journalists in Hong Kong [6]. Snowden began his career at the CIA in May 2006, where he was assigned to the global communications division in Langley, Virginia. When it became clear to the CIA that Snowden was at the top of his field, he was sent to be trained full-time as a technology specialist. After completing the six-month training program, he was sent to Geneva, Switzerland under diplomatic cover, where he undertook his first foreign mission as the employee responsible for computer network security [7] [8].

## A. PRISM and Upstream

After working in Switzerland [9] for the CIA and in Japan [8] for the NSA, Snowden moved to employment in the private sector. In 2009, at the consulting firm Booz Allen Hamilton, he was assigned to be a system administrator for an NSA office in Hawaii [10]. There, Snowden had access to top-secret, classified data including the surveillance program PRISM, which was directly connected to the servers of Google, Microsoft, Facebook, Apple, Skype, and other major companies. PRISM was used to extract data including text from e-mails and documents, video, and audio files, which helped the NSA and United Kingdom's Government Communications Headquarters (GCHQ) to collect and track targets. Some of the information released by Snowden described how the PRISM system worked. First, when a NSA analyst wanted to add a new surveillance target, they had to present a request to a supervisor, who then had to 'endorse the analyst's reasonable belief'. The identified target also had to be 'a foreign national who was overseas at the time of collection'. In the PRISM Tasking Process, the Federal Bureau of Investigation (FBI) then retrieved information matching the target from data providers (PRISM Provider). After collection, the data was 'processed and analysed' by systems such as PRINTAURA (traffic flow automation), SCISSORS (sorting data types for analysis), NUCLEON (voice data analysis), and PINWALE (video data analysis). To filter the information being input, the tools FALLOUT and CONVEYANCE were used. In the document named 'PRISM Case Notations', which shows how a chosen target was registered. It was revealed that each surveyed target had an individual case notation that was a combination of letters and numbers such as 'P2 E SQC 12 0001234'. The first two identifiers referred to the "PRISM Provider' mentioned earlier – in this example, 'P2' represents Yahoo. The third identifier' represented the Content Type – such as 'E' for e-mail, VoIP, IM (chat) and 'J' for videos, among others. The next three identifiers (in this example, 'SQC') were a fixed code 'denoting the PRISM source collection'. The next two numbers indicated the year when the case notation for the selector was established. Finally, the last seven digits were reserved for the case serial number [11]. The first provider that began PRISM collection was Microsoft in 2007, followed by Yahoo in 2008, Google and Facebook in 2009 and finally Apple in 2012. To search the collected data, an NSA analyst could search the PRISM counterterrorism database which, as of April 5th, 2013, contained 117,675 surveillance targets [11] [12]. Moreover, the counterpart of PRISM - known as 'Upstream' - was even more invasive, according to Snowden. With Upstream, the NSA was able to retrieve data directly from routers, switches, satellites and undersea fibre-optic cables in North America [13] [14] [15] [11].

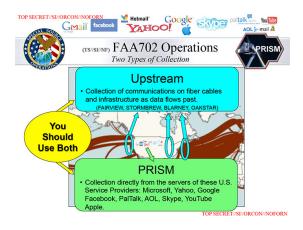


Figure 1. PRISM and Upstream [11]

By using PRISM for stored data and Upstream for live data, the NSA had easy access to the digital information of the world. As shown in Figure 1, Upstream was divided into programs such as FAIRVIEW, BLARNEY, STORMBREW and OAKSTAR. Each surveillance-program had its specialty: FAIRVIEW and STORMBREW were used to collect data about messages and phone calls; BLARNEY accessed internet data through arrangements with companies and internet service providers (ISP); OAKSTAR was used specifically for gathering information outside of the US [16]. For more details about PRISM, The Washington Post discusses the PRISM data

collection program in-depth [12].

#### B. XKeyscore

In addition to PRISM, the NSA had XKeyscore, a software program used to search individuals created by Science Applications International Corporation (SAIC, now known as Leidos). The documents Snowden leaked to journalists revealed that XKeyscore could be used to search 'nearly everything a user does on the Internet' [17]. As shown in Figure 2, every user session was directly processed, analysed and stored in the database. In the documents provided by The Guardian, the NSA explains that XKeyscore provides 'real-time target activity'; thus, each typed character can immediately be seen by the analyst [18].

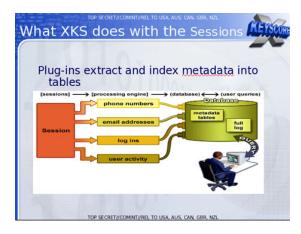


Figure 2. XKeyscore [18]

Since XKeyscore generated vast quantities of data, the NSA had more than 500 servers around the world. When the existing servers were no longer sufficient for the program, they simply added a new server to the system. With the sheer volume of data, the analysts could not search for random data and hope to find something valuable about their target. Instead, they needed to search for strong-selectors such as anomalous events; these events included features such as 'speaks German but is located in Pakistan' or 'using encryption'. Thus, analysts were effectively able to search for everything from names to IP addresses to the language of the individual's internet activity. Snowden reported that NSA analysts are 'able to read any individual's emails' [17] using simply their target's e-mail address and a justification for the search. Given the large volume of data, the content was retained for three to five days and the retrieved metadata was stored for 30 days [19]. However, despite this vast quantity of information, it was revealed that only approximately 300 terrorists were captured as a result of XKeyscore by 2008 [20] [18].

## C. Boundless Informant

While other NSA programs obtained raw metadata, the 'Boundless Informant' tool helped the NSA to track the amount of data being collected from each country. Boundless Informant focused on categorizing information rather than identifying the content of particular messages. The released

documents show that over a 30-day period around March 2013, the NSA collected over 97 billion data entries. With over 14 billion collected data entries, Iran was the most surveyed state, followed by Pakistan with 13.5 billion collected entries. For comparison, there were only 3 billion collected data entries from the US. This provides a sense of the volume of data collected by the NSA in just one month in 2013 [21].

### D. Snowden's Legacy

The Snowden scandal began after The Guardian first published the leaked information provided by Snowden with the article 'NSA collecting phone records of millions of Verizon customers daily' by Glenn Greenwald, one of Snowden's main contact in Hong Kong [22]. Shortly thereafter, The Guardian and The Washington Post published articles revealing PRISM, Boundless Informant and that the US was spying on its European allies. A detailed discussion of these leaks is presented in The Guardian's article 'Edward Snowden and the NSA files - timeline' by Mirren Gidda [23]. Furthermore, all articles published regarding the leaks are available at the Snowden Archive by the Canadian Journalists For Free Expression (CJFE) [24].

#### III. PANAMA PAPERS

'The Panama Papers' refer to documents that were leaked by John Doe, an anonymous whistle-blower who leaked over 11.5 million documents to the German newspaper 'Süddeutsche Zeitung' in early 2015. For comparison, the 'Offshore Leaks' from April 2013 were approximately 10% of the total Panama Papers leak by John Doe [25]. The Panama Papers were revealed with the help of journalists from 80 countries working through the International Consortium of Investigative Journalists (ICIJ) [26]. In April 2016, it was revealed that Mossack Fonseca, a law firm established in Panama in 1977, offered an offshore service to wealthy individuals and entities. In the Panama Papers, the names of more than 140 political officials, celebrities, professional athletes and billionaires were revealed [27].

# A. Preparation to Publish

Due to the large number of documents, the journalists were required to index each document and used optical character recognition to search for text in images, contracts and identity cards [28] [26]. According to ICIJ, the investigative journalists used tools such as I-Hub, Blacklight and Linkurious. I-Hub is described by its users as a 'private Facebook' to communicate safely on sensitive subjects. Blacklight was used to search databases using keywords. Finally, Linkurious was used to visualize and connect data from the searches to uncover connections a human may not identify [29].

### B. Leaked Data

According to the Süddeutsche Zeitung, the Panama Papers cover data from the 1970s to 2016. The law firm created a folder for each shell corporation that contained several data types (Figure 3).

#### The structure of the leak

The 11,5 millionen contain the following file types

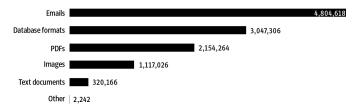


Figure 3. Panama Papers - types of documents [26]

According the ICIJ, the firm created 214,000 offshore companies for individuals in more than 200 countries and territories [30]. While offshore business entities are legal, some were used for illegal activities including fraud and tax evasion. With the complex structure of these shell companies, it's difficult to find the real owner of the company. The Guardian states that more than 100,000 companies created by Mossack Fonseca were held in the British Virgin Islands. The other half was situated in tax havens such as the Bahamas, Panama and the Seychelles. Instead of working face to face with the clients, the law firm used third parties like banks and lawyers mainly from Switzerland, Hong Kong and Panama. China, Hong Kong and Russia are at the top of the list, which have the highest money flow that goes to the offshore companies [31]. The leak revealed that Wladimir Putin, the Russian president, had ties to Mossack Fonseca through a friend named Sergej Roldugin who controls with other friends about \$2 billion worth of assets in the offshore business. Also, the footballer Lionel Messi was named and fined €1.7 million [32]. Gabriel Zucman states that at least \$7.6 trillion, or about 8% of the world's finances, are held in offshore companies [33]. Due to the leaks, the German banks Commerzbank, HSH Nordbank and the Hypovereinsbank were fined approximately €20 million each for their business activities with Mossack Fonseca [26]. According to the ICIJ, the leak helped the United Kingdom to bring back a total of \$252 million, Australia \$92 million, Belgium \$18 million and Germany \$183 million. With all other countries combined, the leak generated over \$1.2 billion in fines and taxes [34]. After two statements in March 2016 in which the firm states that they operated in a legal way [35], the law firm closed in 2018 following the release of the documents

# IV. VAULT 7

On March 7, 2013, WikiLeaks published 24 documents about Vault 7. With over 8,761 documents extracted from the CIA's Center for Cyber Intelligence in Langley, Virginia, WikiLeaks disclosed the agency's secret, worldwide hacking program against operating systems including Windows, Android and iOS, as well as early Samsung smart TVs. The program was named 'Year Zero' and investigated the use of zero day bugs – unknown flaws in the code of an operating system. With Vault 7, WikiLeaks revealed how the CIA used these flaws to hack smartphones and other electronic devices.

The first published set of documents shows that the CIA had a hacking base in the US Consulate in Frankfurt, Germany, which was used to initiate hacking attacks on Europe, China and the Middle East [5].

#### A. WikiLeaks

Since its creation by journalist Julian Assange, WikiLeaks has published more than 10 million classified documents. The goal of WikiLeaks is 'to bring important news and information to the public' [37]. After the first big leak ('Collateral Murder') [3], WikiLeaks published numerous revelations over the years including the Afghanistan War Diary and the Iraq War Logs [38], the Cablegate [39] and Vault 7 [5].

#### B. CIA malware

The documents revealed that the CIA was able to obtain GPS, SMS and audio data from a malware-infected smartphone. The malware also made it possible to turn on the camera and microphone of the device without the user knowing. According to Wikileaks, the CIA had 24 'weaponized' Android zero-day programs that were used to bypass the encryption systems of chat apps like WhatsApp, Signal and Telegram by retrieving the data before the device encrypts the messages. According to WikiLeaks, over 1.15 billion Android phones were sold in 2016 representing a market share of approximately 85%. However, the CIA's malware also targeted Microsoft Windows, Apples iOS and OS-X and the opensource operating system Linux. There were several malware tools, including 'Hammer Drill', 'Brutal Kangaroo', 'Assassin' and 'Medusa', that were used by the CIA to infect devices. The malware named 'Weeping Angel' infected Samsung smart TVs to put the TV in a 'fake' off-mode after being turned off by the user; this turned on the TV's microphone, which is normally shut down after turning the TV off. With the microphone on, private conversations could be recorded and uploaded to the CIA though a Wi-Fi signal. In addition, approximately 300 internet switches from the American technology enterprise Cisco Systems, Inc. were able to be compromised by the CIA [5]. As a result of the leaks, Cisco was able to find and patch the zero-day vulnerability and warn its users [40] [41]. According to an internal audit from the CIA, Vault 7 leaked 91 of the more than 500 malware tools, which shows that the CIA had a wide range of possibilities to infect devices [42].

# V. SUMMARY AND OUTLOOK

In reviewing the leaks described above, it can be seen that everything connected to the World Wide Web can be tracked, analysed and used against individuals for a variety of purposes. However, data leaks can also reveal illegal activities to the public with vast amounts of evidence, as in the case of the Panama Papers. Knowing that intelligence agencies can easily track our digital lives at all times should be alarming. Whereas social media apps such as Facebook and Twitter are now mostly regulated within the European Union by the 'General Data Protection Regulation', apps from outside the regulation, such as the Chinese social media app TikTok, should be

frightening to its users. Thus, to guarantee privacy, offline alternatives that cannot be tracked and analysed should be used.

#### REFERENCES

- [1] "Meaning of leak in English." [Online]. Available: https://dictionary.cambridge.org/dictionary/english/leak
- [2] "Meaning of leak in English." [Online]. Available: https://ahdictionary.com/word/search.html?q=leak
- [3] WikiLeaks, "Collateral Murder," 2010. [Online]. Available: https://collateralmurder.wikileaks.org/
- [4] "The Panama Papers," Süddeutsche Zeitung, 2016. [Online]. Available: https://panamapapers.sueddeutsche.de/en/
- [5] WikiLeaks, "Vault 7," 2017. [Online]. Available: https://wikileaks.org/ciav7p1/
- [6] D. Ignatius, "Edward Snowden took less than previously thought, says James Clapper," *The Washington Post*, 2014. [Online]. Available: https://www.washingtonpost.com/opinions/edward-snowden-took-less-than-previously-thought-says-james-clapper/2014/06/05/054cb9f2-ecee-11e3-93d2-edd4be1f5d9e\_story.html
- [7] M. Memmot, "Snowden Geneva." [Online]. Available: https://www.npr.org/sections/thetwo-way/2013/06/10/190293209/who-is-edward-snowden-the-nsa-leaker?t=1595263634416
- [8] J. Bamford, "The most wanted man in the world," Aug. 2014. [Online]. Available: https://www.wired.com/2014/08/edward-snowden/
- "Snowden's [9] S. Bradley, memoir: Key SWI. 2019 on his time in Geneva," [On-Available: https://www.swissinfo.ch/eng/permanent-record\_ line]. key-takeaways-from-snowden-s-book-on-his-geneva-spying/45239584
- [10] Glenn Greenwald and Ewen MacAskill and Laura Poitras, "Edward Snowden: the whistleblower behind the NSA surveillance revelations," 2013. [Online]. Available: https://www.theguardian.com/world/2013/ jun/09/edward-snowden-nsa-whistleblower-surveillance?CMP=twt\_gu
- [11] "NSA PRISM program slides," The Guardian, 2013. [Online]. Available: https://www.theguardian.com/world/interactive/2013/nov/01/ prism-slides-nsa-document
- [12] "NSA slides explain the PRISM data-collection program," The Washington Post, 2013. [Online]. Available: https://www.washingtonpost. com/wp-srv/special/politics/prism-collection-documents/
- [13] "Upstream vs. PRISM," *Electronic Frontier Foundation*, 2017. [Online]. Available: https://www.eff.org/de/pages/upstream-prism
- [14] J. Ball, "NSA's PRISM surveillance program: how it works and what it can do," *The Guardian*, 2013. [Online]. Available: https://www.theguardian.com/world/2013/jun/08/ nsa-prism-server-collection-facebook-google
- [15] S. Ackerman, "NSA concedes violating surveillance limits and pledges curbs on US email collection," *The Guardian*, 2017. [Online]. Available: https://www.theguardian.com/us-news/2017/apr/28/ nsa-stops-surveillance-us-residents-foreign-targets
- [16] P/K, "Slides about NSA's Upstream collection," Feb. 2014. [Online]. Available: https://www.electrospaces.net/2014/01/ slides-about-nsas-upstream-collection.html
- [17] G. Greenwald, "XKEYSCORE: NSA tool collects 'nearly everything a user does on the internet'," *The Guardian*, 2013. [Online]. Available: https://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data
- [18] "XKEYSCORE presentation from 2008," The Guardian, 2013. [Online]. Available: https://www.theguardian.com/world/interactive/2013/jul/31/nsa-xkeyscore-program-full-presentation
- [19] "3 slides about the XKEYSCORE program," 2014. [Online]. Available: https://www.documentcloud.org/documents/ 894406-nsa-slides-xkeyscore.html
- [20] M. Marquis-Boire, G. Greenwald, and M. Lee, "XKEYSCORE NSA's Google for the World's Private Communications," *The Intercept*, 2015. [Online]. Available: https://theintercept.com/2015/07/01/nsas-google-worlds-private-communications/
- [21] G. Greenwald and E. MacAskill, "Boundless Informant: the NSA's secret tool to track global surveillance data," Jun. 2013. [Online]. Available: https://www.theguardian.com/world/2013/jun/08/ nsa-boundless-informant-global-datamining

- [22] G. Greenwald, "Verizon forced to hand over telephone data," The Guardian, 2013, [Accessed on July 17,2020]. [Online]. Available: https://www.theguardian.com/world/interactive/2013/jun/06/verizon-telephone-data-court-order
- [23] M. Gidda, "Edward Snowden and the NSA files timeline," *The Guardian*, 2013. [Online]. Available: https://www.theguardian.com/world/2013/jun/23/edward-snowden-nsa-files-timeline
- [24] L. Tribe and A. Clement, "Snowden surveillance archive," Canadian Journalists for Free Expression. [Online]. Available: https://snowdenarchive.cjfe.org/greenstone/cgi-bin/library.cgi
- [25] "Offshore leaks," International Consortium of Investigative Journalists, 2013. [Online]. Available: https://www.icij.org/investigations/offshore/
- [26] F. Obermaier, B. Obermayer, V. Wormer, W. Jaschensky, "About the Panama Papers," Süddeutsche Zeitung, 2016. [Online]. Available: https://panamapapers.sueddeutsche.de/articles/56febff0a1bb8d3c3495adf4/
- [27] J. Garside, H. Watt, and D. Pegg, "The Panama Papers: how the world's rich and famous hide their money offshore," *The Guardian*, 2016. [Online]. Available: https://www.theguardian.com/news/2016/ apr/03/the-panama-papers-how-the-worlds-rich-and-famous-hide\ -their-money-offshore
- [28] T. Brewster, "From Encrypted Drives To Amazon's Cloud The Amazing Flight Of The Panama Papers," Forbes, 2016. [Online]. Available: https://www.forbes.com/sites/thomasbrewster/2016/04/05/panama-papers-amazon-encryption-epic-leak/#70139bfe3a34
- [29] P. Romera and C. S. Gallego, "How ICIJ deals with massive data leaks like the Panama Papers and Paradise Papers," *International Consortium of Investigative Journalists*, 2018. [Online]. Available: https://www.icij.org/inside-icij/2018/07/how-icij-deals-with-massive-data\-leaks-like-the-panama-papers-and-paradise-papers/
- [30] W. Kenton, "The Panama Papers: What You Should Know," Investopedia, 2020. [Online]. Available: https://www.investopedia.com/ terms/p/panama-papers.asp
- [31] L. Harding, "What are the Panama Papers? A guide to history's biggest data leak," *The Guardian*, 2016. [Online]. Available: https://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-panama-papers
- [32] "Messi, father guilty of tax fraud, given suspended 21-month sentences," Sports Illustrated, 2016. [Online]. Available: https://www.si.com/soccer/2016/07/06/lionel-messi-tax-fraud-prison-jail-sentence
- [33] G. Zucman, "The hidden wealth of nations," 2015. [Online]. Available: https://press.uchicago.edu/ucp/books/book/chicago/H/bo20159822.html
- [34] D. Dalby and A. Wilson-Chapman, "Panama Papers helps recover more than 1.2 billion around the world," International Consortium of Investigative Journalists, 2019. [Online]. Available: https://www.icij.org/investigations/panama-papers/ panama-papers-helps-recover-more-than-1-2-billion-around-the-world/
- [35] "The official statement of Mossack Fonseca," Süddeutsche Zeitung, 2016. [Online]. Available: https://panamapapers.sueddeutsche. de/articles/56febfc0a1bb8d3c3495adf0/
- [36] N. Slawson, "Mossack Fonseca law firm to shut down after Panama Papers tax scandal," *The Guardian*, 2018. [Online]. Available: https://www.theguardian.com/world/2018/mar/14/ mossack-fonseca-shut-down-panama-papers
- [37] "About What is WikiLeaks?" WikiLeaks, 2011. [Online]. Available: https://wikileaks.org/About.html
- [38] "WarDiaries Iraq and Afghan War Diaries Explorer," WikiLeaks, 2010. [Online]. Available: https://wardiary.wikileaks.org/
- [39] "Cablegate: 250,000 US Embassy Diplomatic Cables," WikiLeaks, 2010. [Online]. Available: https://www.wikileaks.org/Cablegate-250-000-US-Embassy.html
- [40] E. Kovacs, "Cisco Finds Zero-Day Vulnerability in 'Vault 7' Leak," SecurityWeek, 2017. [Online]. Available: https://www.securityweek.com/cisco-finds-zero-day-vulnerability-vault-7-leak
- [41] O. Santos, "The WikiLeaks Vault 7 Leak What We Know So Far," *Cisco*, 2017. [Online]. Available: https://blogs.cisco.com/security/the-wikileaks-vault-7-leak-what-we-know-so-far
- [42] United States District Court Southern District of New York, "USA vs Joshua Adam Schulte," 2020. [Online]. Available: https://www. documentcloud.org/documents/6771808-20200206-REDACTED.html

All online sources and references were last accessed on July 24, 2020.