Cybercrime Profiling:

An Analysis of Correlation between Web Hacking Perpetrators and Victims

Gaberial N. Campese, M.S.

Dr. Nima Zahadat, Ph.D.

The George Washington University

Abstract

The proliferation of cybercrime, especially web hacking, continues to be of growing concern for any computer user around the globe. Because of drastic impacts on both nations and enterprises, it is crucial for investigators to respond by profiling the actors in the cybercrime space. Several articles describe the effects and characteristics of cybercriminals, yet fail to specifically detail the means by which web attacks are committed. In addition, responses and impacts concerning cyberattacks are among consistent subjects of investigation. Further research will provide deeper insight on targets of web hacking syndicates and where they operate with their malicious intents. Python is used for the purpose of data mining with a massive dataset containing 212,093 instances of web hacking between the years of 1998 and 2015. Web hacking is of scope due to commercial reliance on the internet. Correlation between cybercrime networks, victims, and countries of occurrence will be revealed.

Keywords

web hacking, cybercrime, countries, correlation, cybercriminals, profiling, data mining, trends

Introduction

In response to the growing reliance on technology to interact globally, numerous cybersecurity trends and vulnerabilities for individuals are observed. As Ajayi (2016) urges, “It is apparent that the telecommunications medium is the most economical way to transact businesses irrespective of the geographical separations between trading partners”. Utilizing technology as a medium of exchange poses vast reward for cybercriminals no matter the boundaries making this a serious initiative for e-commerce. Web hacking in particular infringes

upon the company's ability to conduct business. The unfortunate prominence of cyberattacks increases detriment among countries and businesses with the rippling of loss incurred by e-commerce seen in the economy. Cybercrime case reports for China show that 68.8% of targets are commercial systems, similar to the United States (Andreasson, 2011). This substantial percentage of targets within countries attract the attention of government entities because of concern for themselves and the well-being of the nation as a whole. The United States government possesses a large amount of sensitive data, making it a goldmine for a cyberhacker. With the possibility to be attacked from anywhere in the world, major concerns for the public sector revolve around the ability to use the internet for espionage and terrorism (Sarre, Lau, & Chang, 2018).

With such negative impacts, several key responses to cybercrime are enacted by all organizations. The actions taken against cybercrime are often times unfavorable by some or all, yet necessary. For instance, the United States Security Exchange Commission has made it mandatory for firms to report an instance of being hacked with the 2011 Disclosure Guidance (SEC, 2011). The law intensifies a company's pressure to avoid cyberattacks in that a disclosure of being attacked could easily contribute to harming reputation and revenue, but the consumer deserves the liberty to know whether their personal data is at risk. The costs to fight this epidemic can be seen globally with average cybersecurity spending highest in German companies at $298,258 and lowest being among Australian companies at $106,904 ("Cybercrime Impacts", 2011). The response by law enforcement is seen with the development of teams primarily for investigations of cybercrime and terrorism. To reiterate, the allocation of extensive resources will always be preferably avoidable by any entity seeking to be profitable.

Understanding the profile of cybercrime networks is crucial in the combatting of cyberattack exploits. The need to recognize the behavior and attributes of cybercrime organizations intensifies with the fact that crimes of this nature are sometimes unable to be solved by law enforcement (Sarre, Lau, & Chang, 2018). Data mining the web hacking dataset with Python works to identify top cybercrime groups while describing key variable relationships including dates, victims, web servers, and operating systems of choice.

Literature Review

Each article provides a unique perspective on the attributes of cybercriminals. Broadhurst, Grabosky, Alazab, and Chon (2014) investigate the nature of groups partaking in cybercrime by categorizing them as organized, loose, and even state sponsored. The authors imply a broad definition due to the fact that they discover that a group could be highly structured for longevity, while another may be disbanded after their mission is completed. Often hidden from the public, an increase in cybercrime can be connected to the growth in cyberattacks committed by government agencies themselves (Broadhurst et al., 2014).

Leukfeldt, Kleemans, and Stol (2017) map networks based on functions and argue that there are dependency relationships. Core members are said to be the directors of the attacks and sit at the top of the hierarchy. There can also be a hierarchy within core members. Next come professional enablers, who work on their own accord and are responsible for developing malicious services for core members. Recruited enablers align under professional enablers and their distinction is made clear:

> "Recruited enablers also provide services to the core members, but they are encouraged or forced by the core members to do this. They have access to information that is of interest to the core members or they are able to provide 'simple' services; services that

> core members could also perform on their own or without which the crime script could still be executed" (Leukfeldt et al., 2017).

Finally, money mules rank at the bottom and have the responsibility of holding stolen money acquired through cyberattacks. The money is then cashed by any one of the personas in the network to avoid creating a money trail (Leukfeldt et al., 2017). This appears to be a fairly complex, articulate cybercrime network topology.

Similar to enablers in the previous literature, there are instances of hackers having their own business of selling malicious software to cybercriminals (Cai, Du, Xin, & Chang, 2018). Cai et al. (2018) narrowed their scope to China's cybercrime challenges arising from the propagation of internet usage. The study produced their findings by analyzing 448 Chinese sentencing documents based on real asset theft, virtual asset theft, online frauds, and stolen accounts. Since more than one group can be using redistributed tools, it may be useful in identifying cybercriminals if there are similarities.

The next series of research articles describe internet and web hacking more specifically. Agent-Based Modeling is used to assess internet hacking by modeling an individual's inclination to hack based on their agent classification as a user, hacker, or inciter (Tang, Bagchi, & Jain, 2009). A positive correlation is found between technical expertise and inclination of hacking. The following are in order from least to greatest in the correlation of draw to hacking and technical expertise: users, hackers, and inciters. The user is considered to be any internet user, while inciters are the elite (Tang et al., 2009). It is important to highlight that individuals can fluctuate on the spectrum. For example, a user can eventually grow into a hacker with continued internet usage. Enforcers are another agent identified as "the counterforce to keep hacking in check" (Tang et al., 2009).

The final literature reviewed is more concise and specific than the previous articles. Han et al. (2017) analyzes web hacking through profiling an attack committed by North Korea against South Korea in attempts of defacement. The article emphasizes web hacking profiling can be extremely advantageous in investigations by law enforcement because they are able to relate the North Korean incidents to Sony cyberattacks with a close, calculated similarity score. Note that the author's scope strictly focuses on the relationship between the North Korean and Sony attacks, yet it does an outstanding job at illustrating the effectiveness of a web hacking statistical analysis. The authors were able to do a brief, statistical analysis through the construction of their own web hacking dataset. Moreover, the article claims:

> "Intelligence derived from WHAP cannot guarantee a full accuracy, but highlights hidden similarities between various cases and assists investigator in timely decision-making" (Han et al., 2017).

The authors have since made the dataset publicly available for researchers to uncover further insights and trends in cybercrime.

Research Methodology

It is evident, that the turmoil inflicted upon organizations and overall structure is often described in detail; however, studies fall short in identifying the particulars surrounding the manners by which cybercrime units commit web attacks. Even responses to cyberattacks are grazed by researchers, yet a more formulated plan to resist can be produced by appropriately profiling cybercrime wrongdoers. The data mining research in Python works to correlate relationships between cybercriminals and victims by grouping the web servers and operating systems used by victims in web hacking attacks. The association of these key variables can then be made to assess patterns amongst deviant actors within particular dates and locations of

manifestation. The victim's web servers, operating systems, and countries can help provide patterns among targets. This can contribute to enhanced relationships between attackers and details in web hacking attacks. With this analysis, investigators could potentially narrow down and pursue committers of web hacking.

Data

The dataset was obtained from the Hacking and Countermeasure Research Lab in CSV format. The sheet contains 212,093 instances of attacks between 1998-2015 represented as rows along with important details represented as columns including the Date, URL, IP, Notify, Country, OS, Web Server, Encoding, and Language. It is important to note that the data ends in September 2015 and that the OS and web servers represent those used by the victims. There were some inconsistencies with the entry of the countries, mainly with capitalization, which were mitigated with Excel. The dataset was read into Python as a CSV file and the Notify column was renamed to Attacker. The URL, IP, encoding, and language columns were dropped using Python. Finally, Year and Month columns were generated by cleaning the Date column in Python.

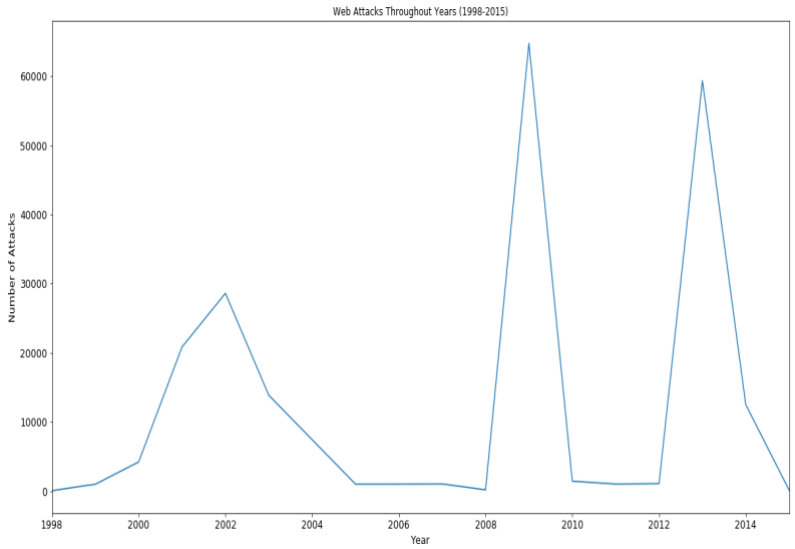
Data Analysis

The data analysis was conducted using Python in Jupyter Notebook using pandas and matplotlib libraries. The research took a funneled approach, in that it went from broad to more narrow details. The specific columns analyzed were Year, Month, OS, Web Server, Attacker, and Country. Each column was mined specifically through grouping and counting methods. Since every attack had a date, the Date column was renamed to Number of Attacks within the grouping methods. For instance, Python was used to group the Number of Attacks by Attacker

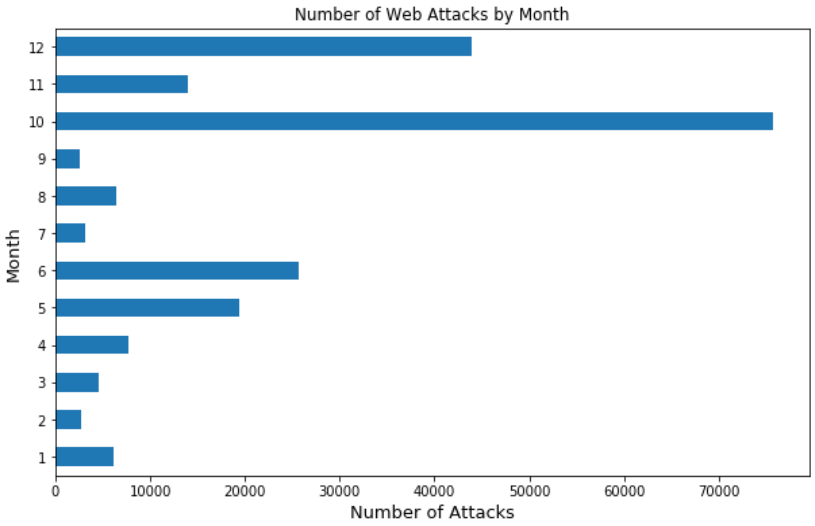
returning the number of attacks committed by each perpetrator. Several types of plots were created depending on the use case. Finally, three methods were developed that allow input of any web hacking group and return the top targeted countries, web servers, or operating systems for that specific perpetrator.

Key Findings

The key findings discovered mainly pertain to the attacker, victim OS and web server, dates, and countries of occurrence. Starting with a historical overview, it can be seen in Graphic 1 that 2009 had the most web hacking attacks followed by 2013. Graphic 2 depicts October as the top month of occurrence by a significant margin followed by December. Recall, the dataset cuts off in September 2015, which may be a contributing factor to 2015 being lower since October and December are excluded.

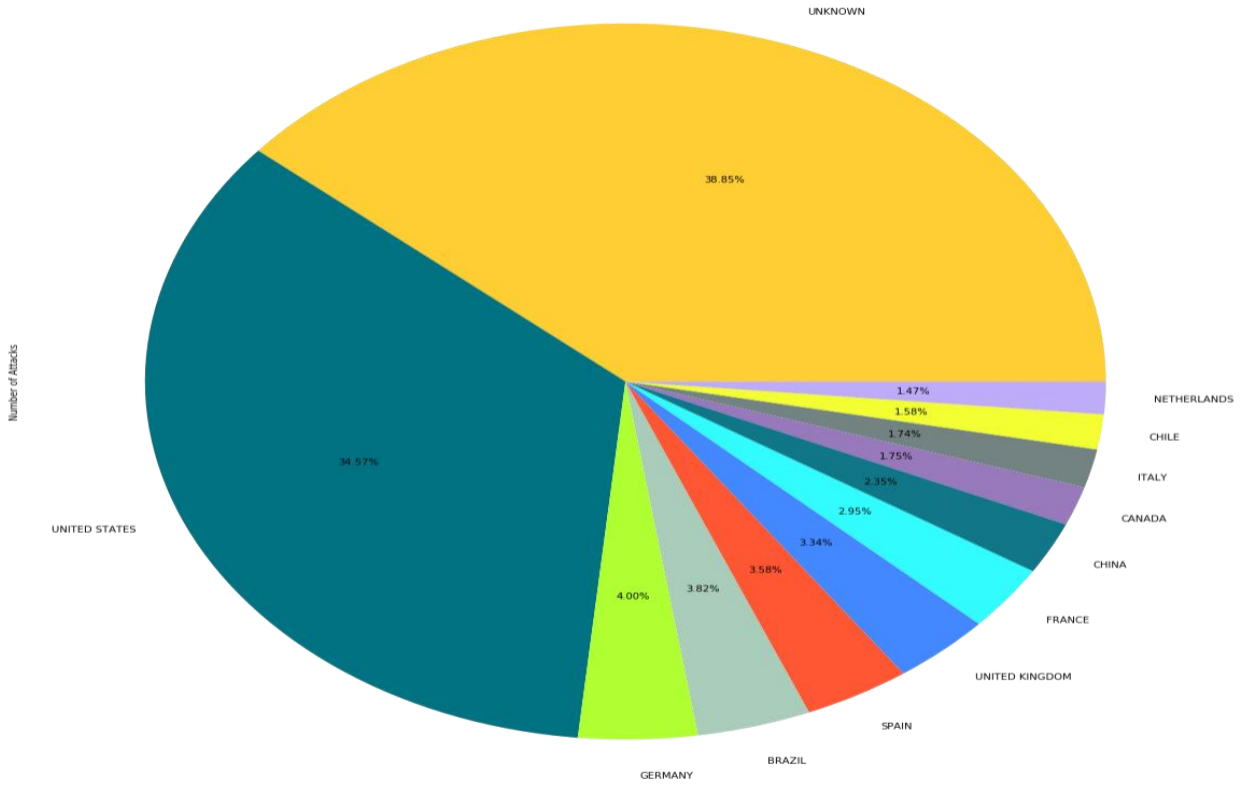


Graphic 1

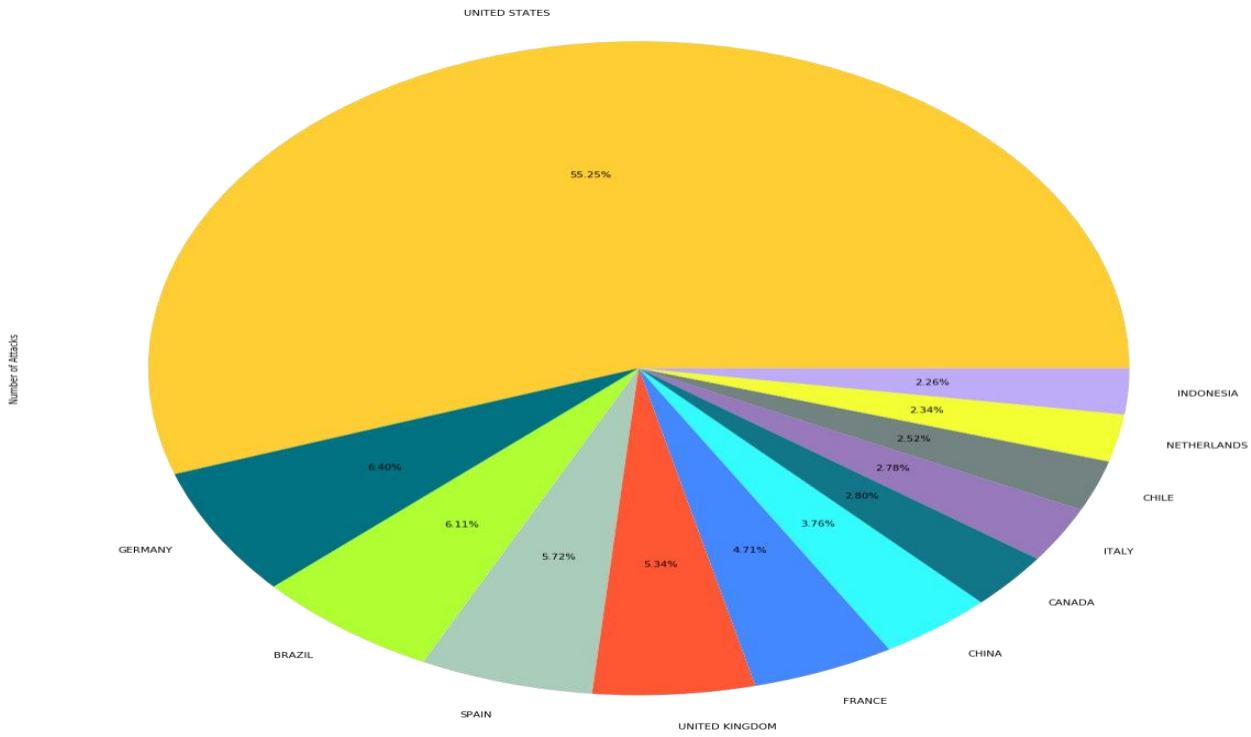


Graphic 2

The top twelve countries of web hacking occurrences, in order, were found to be Unknown, United States, Germany, Brazil, Spain, United Kingdom, France, China, Canada, Italy, Chile, and the Netherlands. Despite the prominence of unknown countries, the margin between the United States and remainder of the countries cannot be ignored in Graphics 3 and 4. The United States accounts for 34.57% of attacks with the presence of unknowns at 38.85% among the top twelve. Without unknowns, the United States hosts 55.25% of web hacking instances within the top twelve countries. The second known country is Germany at a mere 6.4%, making the US a prime home to web hacking victims.

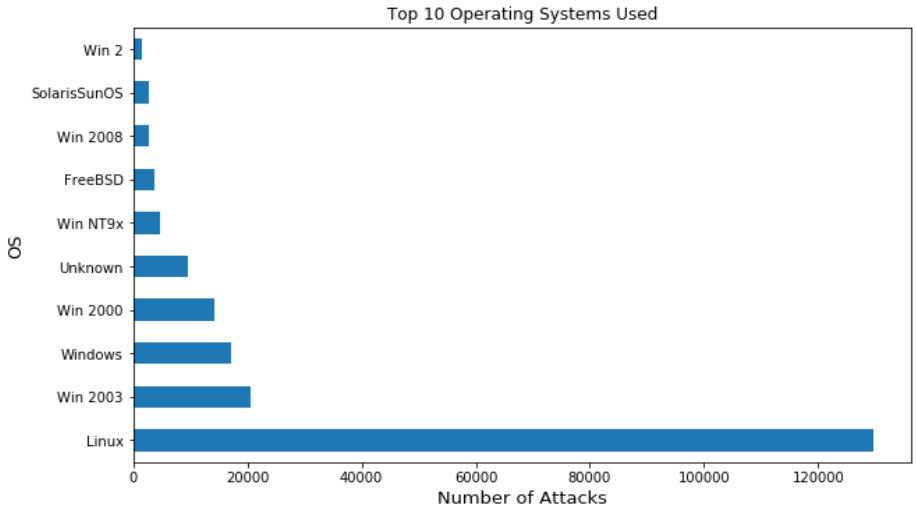


Graphic 3

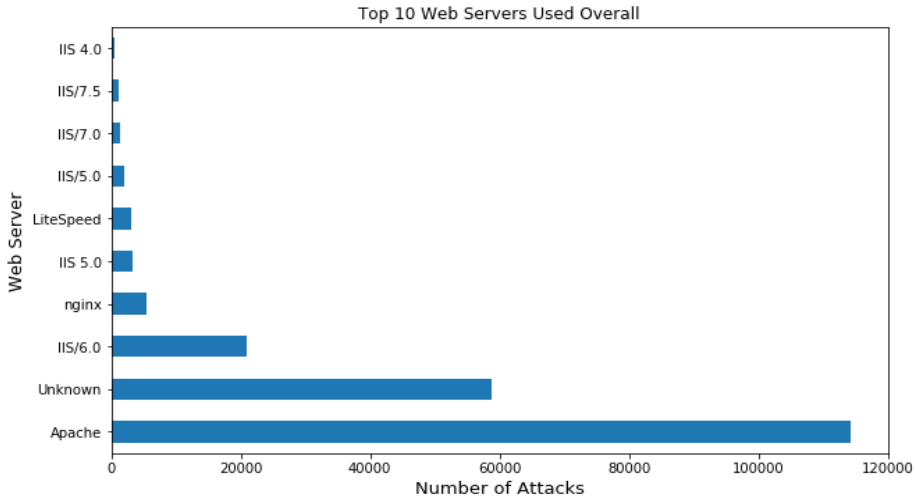


Graphic 4

Moving on to the profile of the targets, their top ten operating systems and web servers are depicted in Graphics 5 and 6. Linux and Apache are the most used OS and web server overall by the victims. Linux accounted for 61.2% of the victim's operating systems and Apache being used at 53.9% for web servers.

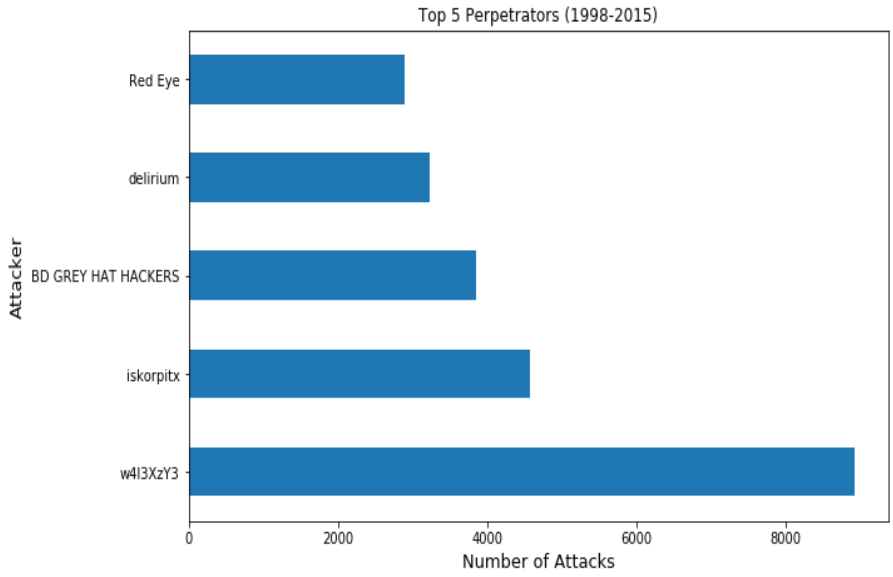


Graphic 5



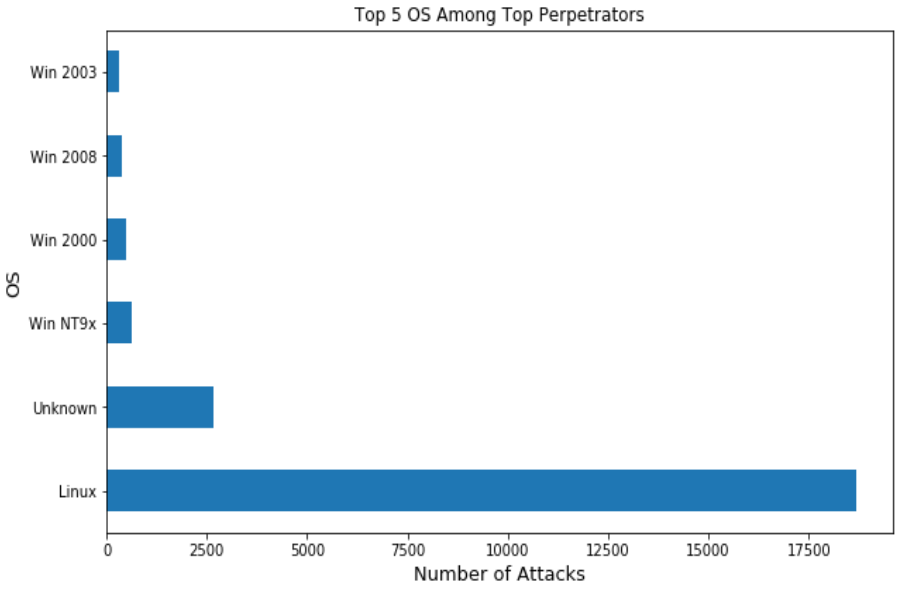
Graphic 6

The top five perpetrators were then identified as w4l3XzY3, iskorpitx, BD GREY HAT HACKERS, delirium, and Red Eye as seen in Graphic 7. These five attackers were responsible for 11.1% of the 212,093 web hacking cases.

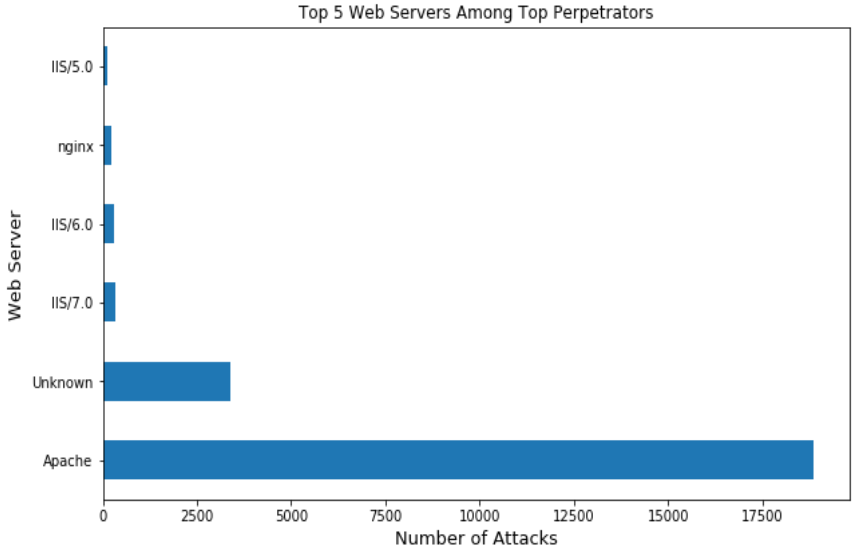


Graphic 7

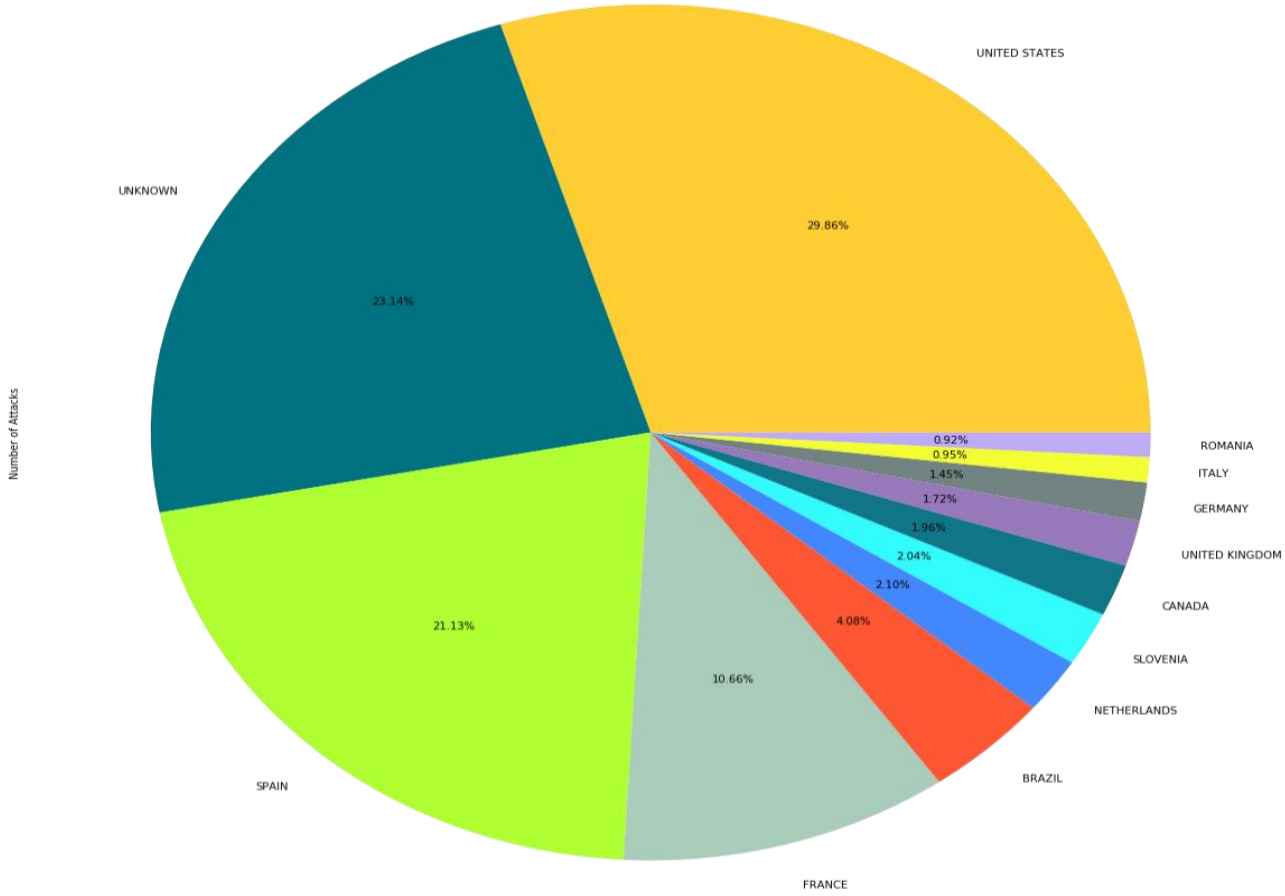
The habits of the top five perpetrators were then analyzed by looking at their targeted operating systems, web servers, and countries. As detailed in Graphics 8-10, the top five perpetrators have similar patterns in comparison to the overall dataset concerning these attributes. 79.5% of their victims were using Linux and 80.3% used Apache. Most of the top five attackers targeted victims in the United States at 29.86%. There are a considerable amount of unknowns at 23.14%. Spain's healthy portion can be attributed to the fact that most of w4l3XzY3's attacks were in Spain.



Graphic 8



Graphic 9



Graphic 10

Through the use of three developed Python functions, each of the top five attackers were profiled based on their targeted Country, OS, and Web Server. w4l3XzY3 primarily focused on Spain, Linux, and Apache victims. Iskorpitx mainly targeted unknown countries, Linux, and Apache victims. BD GREY HAT HACKERS centered their hacks in the United States, Linux, and Apache. Delirium conducted most of their attacks on the United States, unknown operating systems, and unknown web server targets. Finally, Red Eye appeared to target unknown countries, Linux, and unknown web servers. The inconsistences with the existence of unknowns in the dataset pose concern when profiling web hacking perpetrators. For example, delirium's top targeted OS and web server were both found to be unknown. While the research provides us with

some insight on the patterns of these groups, it is important to note that it lacks consistency in providing a vivid portrait for every offender.

Recommendations

The following are recommendations for future researchers that may use the previous study as an asset. To provide a more formulated picture, research conducted in the future should focus on more separable attributes of web hacking cases. Moreover, it is a great idea to expand and focus on other types of cybercrimes. More specifically, future researchers should take advantage of the insight extraction methods offered by data mining with Python and apply it to profiling cybercriminals. Python tools offer powerful methods in effective pattern and trend recognition.

Conclusion

Web hacking does not appear to be slowing down as the world relies heavily on the Internet for e-commerce, research, interacting, and much more. The study recognizes that there is a trend in the United States being a prime country for web hacking cases. The research also indicates a pattern with the overwhelming usage of Linux and Apache among web attack victims; however, this observation does not necessarily mean that users of Linux and Apache are more prone to being attacked. While it is great to see patterns among the web servers and operating systems being targeted, it can be extremely difficult to narrow down the attackers since the vast majority of their targets are using the same OS and Web Server. All in all, profiling cybercriminals committing web attacks is volatile yet ever so important in today's world of technological dependence.

Biographies

Gaberial Campese is pursuing his M.S. in Data Science at the George Washington University during the evening hours. He previously studied Information Systems at the University of Florida's Warrington College of Business.

Dr. Zahadat is the Data Mining professor at the George Washington University. He has held various positions in the industry including Chief Security Officer and Chief Information Officer. His research areas are Data Science, Digital Forensics, Machine Learning, Mobile Security, and Security Policy Management.

References

Ajayi, E. F. G. (2016). The Impact of Cybercrimes on Global Trade and Commerce. *International Journal of Information Security & Cybercrime*, *5*(2), 13-17.

Andreasson, K. J. (2011). K. J. Andreasson (Ed.), *Cybersecurity: public sector threats and responses*. Boca Raton, FL: CRC Press.

Broadhurst, R., Grabosky, P., Alazab, M., & Chon, S. (2014). Organizations and Cyber Crime: An Analysis of the Nature of Groups engaged in Cyber Crime. *International Journal of Cyber Criminology*, *8*(1), 1–20.

Cai, T., Du, L., Xin, Y., & Chang, L. (2018). Characteristics of cybercrimes: evidence from Chinese judgment documents. *Police Practice and Research*, *19:6*, 582-595.

CF Disclosure Guidance, United States SEC Topic No. 2 (2011).

Cybercrime Impacts, Costs Vary by Country, but Motivations Stay the Same. (2012). *Security: Solutions for Enterprise Security Leaders*, *49*(7), 12.

Han, M. L., Han, H. C., Kang, A. R., Kwak, B. I., Mohaisen, A., & Kim, H. K. (2017). WHAP: Web-hacking profiling using Case-Based Reasoning. *2016 IEEE Conference on Communications and Network Security, CNS 2016*, 344-345.

Han, M. L., Han, H. C., Kang, A. R., Kwak, B. I., Mohaisen, A., & Kim, H. K. (2017). *Web-Hacking Dataset*. Retrieved from http://ocslab.hksecurity.net/Datasets/web-hacking-profiling.

Leukfeldt, E., Kleemans, E., & Stol, W. (2017). A typology of cybercriminal networks: from

low-tech all-rounders to high-tech specialists. *Crime, Law & Social Change*, *67*(1), 21–37.

Sarre, R., Lau, L., & Chang, L. (2018). Responding to cybercrime: current trends, *Police Practice and Research*, *19:6*, 515-518.

Tang, Z., Bagchi, K., & Jain, A. (2009). Explorative Assessment of Internet Hacking: An Agent-Based Modeling Approach. *Journal of Information Privacy & Security*, *5*(2), 42–64.

3 Potential Publishing Journals: Journal of Information Privacy and Security, Police Practice and Research, and International Journal of Cyber Criminology.