International Journal of Cybersecurity Intelligence & Cybercrime

Volume 6 | Issue 1 Article 5

3-31-2023

Understanding the Connection Between Hackers and Their Hacks: Analyzing USDOJ Reports for Hacker Profiles

IJCIC Center for Cybercrime Investigation & Cybersecurity

Gerstenfeld. (2023). *International Journal of Cybersecurity Intelligence and Cybercrime, 6*(1), 59-76.

Understanding the Connection Between Hackers and Their Hacks: Analyzing USDOJ Reports for Hacker Profiles

Joshua Gerstenfeld*, B.A., University of Florida, U.S.A.

Keywords: Cybercriminal, Cybersecurity, Cyber Investigations, Profiling

Abstract:
Recently, it seems as if hacking-related stories can be found in the news every day. To study, and hopefully prevent, this new type of crime, the field of cyber criminology has emerged. This study adds to the existing cybercrime literature by examining hacking behavior specifically. It determines if there is a relationship between the age, gender, and nationality of hackers and characteristics of the cyberattacks that they perpetrate. To do this, this study analyzes 122 United States Department of Justice (USDOJ) press reports from January 2019 to December 2021. Some key results include the finding that older hackers and international hackers are more likely to build/maintain software in their cyberattacks. Also, older hackers are less likely to use follow-up access in their hacks compared to younger hackers. Finally, international hackers are more likely to have more sophisticated attacks than nationals, and individual hackers are less likely to have sophisticated hacks than those working in groups. Implications of this study are that law enforcement can create a more accurate profile of hackers based on their hacks to guide them in investigations.

Introduction

For almost all of human history, the idea of a computer, an internet, or even an electronic screen seemed more like fiction than reality. Yet today's technology is very much real, and constantly increasing in use (Greenwood, 2022). The increased use of technology gives many benefits for society, but amongst the negative effects of increased technology use is the corresponding increase in opportunity to commit cybercrime (Miró-Llinares & Moneva, 2019). Although there are many types of cybercrime, arguably the most famous (or infamous) type of cybercrime is hacking. Hacking is an activity that has long fascinated the public, as evidenced by the history of romanticization of it in both film and television (Cliffe). The interest in hacking is understandable, as stories involving this dangerous form of cybercrime appear to be more frequent with each passing day.

Academia is also interested in studying hacking, although research on understanding the cybercriminals who engage in the activity has faced challenges (Hutchings & Holt, 2018). This is for a variety of reasons, not the least of which being that "hacking" is an ambiguous term that means different things to different people (Hutchings & Holt, 2018). Various groups (social scientists, computer scientists, older academics) have inconsistent definitions for the word, which limits the ability of researchers to effectively collaborate on the topic of hackers, since few agree what causes an individual to even be classified as such (Hutchings & Holt, 2018). Even this paragraph's earlier reference to hackers as "cybercriminals" could be contested by many professionals, as "hacking" is also used to refer to legitimate and legal activities, such as penetration

*Corresponding author
Joshua Gerstenfeld. B.A., Department of Sociology and Criminology & Law, University of Florida, Gainesville, FL 32611, U.S.A.
Email: jrgerstenfeld@gmail.com

testing and bug bounties.

The confusion stems from the fact that originally, hackers referred to tech savvy people driven by a "hacker ethic," which prioritized liberal values such as freedom of speech and information sharing (Jaquet-Chiffelle & Loi, 2020). However, as computers and Wi-Fi became more commonplace, hacking became a lucrative endeavor, and the ideological incentives of early hackers were replaced with economic ones (Jaquet-Chiffelle & Loi, 2020). This process led to the modern labels of white hat and black hat hackers (Jaquet-Chiffelle & Loi, 2020), a dichotomous classification system where the "good" white hats legally protect the cyber world from the criminal actions of the "evil" black hats (Chiesa et al., 2018). A third category, gray hats, refer to "morally ambiguous" hackers that refuse to identify with either group (Chiesa et al., 2018). Despite the fact that the hat labels are easy to understand, researchers have argued that they are not sufficient to explain hacking behavior (Jaquet-Chiffelle & Loi, 2020) (Chng et al., 2022).

In order to remedy this issue, Chng et al. (2022) created a classification system of 13 types of hackers: "novices, students, cyberpunks, old guards, insiders, petty thieves, professionals, nation states, hacktivists, digital pirates, online sex offenders, crowdsourcers, and crime facilitators." This designation also ranks hackers by skill, with novices and students being at the lower end of the spectrum, using technologies developed by more experienced hackers, such as professionals and old guards (Chng et al., 2022). As understanding of hacker behavior continues to grow, more classifications will likely be added to this system (Chng et al., 2022). Although these definitions are essential for future research, the constantly changing hacking labels throughout the years have unfortunately caused the definition for "hacker" to become complicated. It is easy to see how such complications could negatively affect research.

Further issues with understanding hacking stem from the fact that cybercriminal offenders, including hackers, are notoriously difficult to access (Hutchings & Holt, 2018). This is due to a combination of factors which include the lack of physical interaction while hacking, the distrustful nature of many cybercrime offenders, and the relatively small number of hackers (Hutchings & Holt, 2018). Furthermore, cybercrime, which encompasses hacking, is inherently an interdisciplinary field (a combination of both computer science and criminology), a fact that can create unique challenges for researchers (Payne & Hadzhidimova, 2020). These difficulties have a practical effect, as law enforcement is negatively affected by the lack of research regarding cyber-offenders (Bachmann, 2010).

These issues highlight the need for new research on hacking, research that has both a technical component, and the capability to investigate hackers without interviewing them. This study aims to accomplish both of these objectives by using open-source USDOJ data regarding hackers and their hacks. The purpose of this study is twofold: primarily, it seeks to find relationships between characteristics of a hack completed by a hacker and personal characteristics of that hacker, in order to aid cybercrime investigators in their investigations. Ideally, this research, and similar research, will allow law enforcement to create a preliminary profile of a hacker based on characteristics of that hacker's hacks. This study furthers that goal by reviewing both technical and non-technical aspects of hacks. Analysis for the profiling aspect of this article was conducted with t tests and chi-squared tests.

A secondary, but related, part of this study is to view both how characteristics of hacks relate to the hacker themselves, and how characteristics of the hackers relate to the attack sophistication of the hack.

This was done with logistics testing, and the aim of these tests was to generate research that determined what relationship, if any, existed between the personal and technical aspects of hacking. Although the secondary aspect of the study has fewer practical consequences than the first, it is this paper's belief that it is important to understand the relationship between who a hacker is and how they hack, in order to create theories regarding what causes individuals to engage and succeed in criminal hacking. Despite the separation between the two objectives, both aspects were closely related and used the same data, which justifies the reporting of the two in the same paper.

Literature Review

Although there have been difficulties with studying hackers, important developments have been made in researching the age, gender, residency, and mental characteristics of these cybercriminals (e.g., Bachmann, 2010; Steinmetz, 2015; Turgeman-Goldschmidt, 2008). Arguably, the most important of these developments are the findings showing that the average hacker is usually white and male (Bachmann, 2010; Steinmetz, 2015; Turgeman-Goldschmidt, 2008), with varying ages (Steinmetz, 2015; Turgeman-Goldschmidt, 2008). This knowledge provides researchers with a basic profile of hackers, which can be used as a springboard for future research.

For example, the gender gap in hacking has been heavily researched, and many have offered possible explanations for this phenomenon (e.g., Holt & Steinmetz, 2020; Donner, 2016). One possible reason for the gender discrepancy is that it is a result of there being a disproportionate number of male computer science majors, who subsequently become programmers, and eventually acquire the skills necessary to commit cybercrime (Grispos, 2019), including hacking. Science Technology Engineering and Math (STEM) fields, which computer science is a member of, have long had a mostly male composition (Charlesworth & Banaji, 2019). Efforts have been made to change this fact, but the disparity between males and females in STEM still exists (Charlesworth & Banaji, 2019), and likely affects the gender ratio of hackers. Other approaches to explain the hacking gender divide have tried to make use of criminological theories. One such study used power-control theory to explain the difference and found that the popular criminological theory could account for some gender variability in hacking (Holt & Steinmetz, 2020).

Regarding research conducted on international hackers, a study on international cyber offenders prosecuted in the United States found that these individuals are usually male, generally come from countries that are political adversaries to the United States, and often work with citizens of other countries in their cybercrimes (Hadzhidimova & Payne, 2019). Although this research was not focused specifically on hacking, a large portion of the data set was based on hacker prosecutions (Hadzhidimova & Payne, 2019), so this data is still useful in understanding international hackers. Other research found that, unsurprisingly, hackers with nationalistic tendencies were more likely to intend to hack enemy nations than hackers with fewer nationalistic tendencies (Woo, 2003). Additionally, a study on cybercrime in South Korea found that Korean hackers were more likely to hack other nations than their own (Back et al., 2019). This research also found that adult hackers were more financially motivated than younger hackers in their cybercrimes (Back et al., 2019).

Further studies that focused on motivations of hackers found that many individuals hack out of a desire for fun, novelty, knowledge, or computer virtuosity (Turgeman-Goldschmidt, 2005). Another source

assigned various motivations for hackers which included ego, exposure, monetary gain, revenge, sabotage, disinformation, and Infowar (cyber-war related) (Shoemaker & Kennedy, 2009). A similar, but more formal, classification scheme was proposed in a recent article which gave seven distinct hacker motivations, being: curiosity, financial, notoriety, revenge, recreation, ideology, and sexual impulses (Chng et al., 2022).

The multitude of hacker motivation categorizations is conducive to further research, since understanding hacker motivation can lead to understanding hacker behavior, as was demonstrated with research about Israeli hackers (Turgeman-Goldschmidt, 2008). This study, conducted as an interview of 54 hackers, found that when hackers peak in their technical abilities, the joy they gain from hacking decreases (Turgeman-Goldschmidt, 2008). The study attributed this decrease in hacking pleasure as a factor in hackers becoming "ex-hackers" (Turgeman-Goldschmidt, 2008). Ex-hackers are those that gave up hacking to take legitimate and lawful jobs in the cybersecurity industry (Turgeman-Goldschmidt, 2008). Interestingly, the interviews found that these "ex-hackers" saw no moral problem with their previous hacking (Turgeman-Goldschmidt, 2008), which may imply that ethical considerations are not a major part of hackers' decisions to leave the criminal life.

A seperate interview of hackers revealed that traits they share in common include a high tolerance for ambiguity, self-rated high creativity and curiosity, and using personal reflection as a means of building or maintaining mental models used to aid their hacking (Summers, 2015). Another study, based on interviews of hackers at a hacking convention, found that compared to the general public hackers perceive themselves to have a more "analytical and rational thinking style" and "display a generally higher confidence" in decision-making abilities (Bachmann, 2010). From these analyses, it is clear that there is a relationship between hacking and certain mental or physical characteristics, but the full extent of this relationship remains to be seen.

Unfortunately, studies to expand on such relationships can be difficult to accomplish, as criminological research questions are often guided by social theories (Bachman et al., 2008). This is an issue for hacking especially, as there is a debate on whether traditional criminological theories can even be applied to the cyber realm (Payne et al., 2018). Despite this fact, there has been some success when applying traditional criminological theories to hacking. For example, a recent study of younger hackers found strong support for self-control theory, and partial support for social bonding theory in an effort to explain juvenile computer hacking (Back et al., 2018). Social learning theory has also been proposed as an explanation for hacking because hackers receive positive reinforcement from the hacker community, which could motivate them to continue engaging in cybercrime (Chng et al., 2022). Furthermore, as previously mentioned, power-control theory can be used to explain some degree of hacking behavior (Holt & Steinmetz, 2020).

There is also some information available about how hacker characteristics affect actual hacks. Bachmann (2010) found that although age and gender do not affect the self-reported success rate of hackers, minority and student hackers report lower success rates than white and non-student hackers do. However, the article was quick to note that since there were few minority and female hackers included in the study (a common problem in hacking studies) more research is needed to confirm these results (Bachmann, 2010). The study also found that the mindset of hackers affects their success rate, by demonstrating that hackers with a preference for analytical thinking and lower risk propensity are more successful than hackers without these values (Bachmann, 2010). This type of research, which integrates understanding of hackers' personal

characteristics with their technical abilities, is necessary to expand cybercrime literature. This article's secondary research objective seeks to further that goal, by examining how the personal characteristics of age, gender, and residency are related to the hacking method a hacker used.

The primary objective of this paper, to provide law enforcement with tools to profile hackers, can be accomplished by studying the same variables. Specifically, this study seeks to determine the likelihood of a hacker's age, gender, and residency, based on characteristics of a hack. This is similar to an idea proposed by Chng et al. (2022), who suggested that it is possible to identify hackers based on "actions in preparation for or during an attack."

Age, gender, and residency were chosen as variables because these characteristics were present and explained in the majority of USDOJ reports (which this paper used as a data set), and because these variables would assist law enforcement in building a profile of hackers. Similarly, the hacking characteristic variables were chosen because they were explained in the reports, and would feasibly be known in many hacking cases. Most of the hacking variables were in some way related to the overall sophistication of the hack, with the exception of two variables: a variable checking if the perpetrator worked alone and a variable checking if the hack had a political/national element. It is expected that hacks with a political/national element will be mostly from international hackers, since prior research demonstrated that nationalistic hackers (who for obvious reasons would be more likely to be involved in hacks with a political/national element) were more likely to hack enemy nations (Woo, 2003). In regards to group hacks, this paper agrees with the prediction of Chng et al. (2022), that attack complexity of a hack can be used as an indicator of whether the hack was committed by an individual or a group.

The goal of this research is to allow law enforcement to apply the profiling tests proposed in this paper to identify hackers. Due to the lack of physical evidence in cybercrimes, information that can be gleaned from hackers' behavioral characteristics is critical to digital forensics investigations (Shoemaker & Kennedy, 2009). Therefore, this study, and studies of a similar nature, have the potential to aid cybercrime investigators in their work.

Methods

To analyze the relationship between the age, gender, and residency of hackers and characteristics of a hack, this paper examined the United States Department of Justice Computer Crime and Intellectual Property Section (USDOJ CCIPS) press releases over a period of three years, from January 2019 to December 2021. Press releases from CCIPS reports provided details of either an arrest, conviction, sentencing, or arraignment of a case that the department was involved with. CCIPS reports dealt with a variety of different cases, but this paper only examined press releases where hacking was used. Furthermore, the term "hack" can apply to many different cybercrimes, but for simplicity and uniformity this paper only classified "unauthorized access to a computer" as a "hack" (despite the fact that other cyberattacks, such as Denial-of-Service Attacks, are sometimes also referred to as hacks) and used the terms "hack" and "hackers" interchangeably with "unauthorized computer access" and "someone who committed unauthorized computer access." Also, juveniles were excluded from the data set.

The CCIPS reports normally contained the perpetrator's age, gender, and residency, which was an important factor in the decision to investigate these variables. Since these variables would be useful in creating a hacker profile, and were present in the data, they were ideal to analyze in this study. As explained in the literature review, research surrounding these characteristics has found both that hackers can be a variety of ages (Steinmetz, 2015; Turgeman-Goldschmidt, 2008) and are rarely female (Bachmann, 2010; Steinmetz, 2015; Turgeman-Goldschmidt, 2008). Regarding international hackers, studies found that nationalistic hackers are more likely to hack enemy nations than their own (Woo, 2003), and that international cyber offenders (a group which includes hackers) prosecuted in the United States usually come from countries that are political adversaries to the United States (Hadzhidimova & Payne, 2019).

In order to examine the relationship between the age, gender, and residency of hackers and hack sophistication, six different "hack characteristic" variables were used. The variables are: the hack was accomplished with insider information, social engineering was used in a hack, the attacker built/maintained software for the hack, the hack was used for follow-up hacks, multiple entities were targeted, and one of the entities targeted included an organization. For the purposes of this paper, social engineering was defined as "tricking the user into revealing computer information or completing a cyber-action." Insider information was defined as "having access to privileged information based on proximity to a target that could be used to aid in a hack." A "follow-up" hack (referred to as follow-up access) was defined as "a hack that could only be completed because of the success of a previous hack committed by the same perpetrator." Although the definitions provided for social engineering, insider information, and follow-up access are not novel interpretations of the terms, the specific wording presented is unique to this study and was generated by the author. The reasoning for the decision to create definitions was so that the study would be able to use precise language that was testable with the information available in the data set, while remaining true to the meanings of the terms.

Additional variables this study checked for included whether the hack had a political/national element and whether the hack was performed by an individual, a pair, or a group. These variables were collected in order to determine a profile of hackers to be used in cyber investigations. However, they are not hack characteristics, so these two variables were not used in determining attack sophistication.

The "insider information" variable was marked as "yes" if the CCIPS report detailed that the perpetrator associated with the targeted person, or, if the target was an organization, worked for/associated with someone who worked for the targeted organization. The "social engineering" variable was marked as "yes" if the article explicitly mentioned social engineering, or if it mentioned that the attacker "tricked the user into revealing computer information or completing a cyber-action." This exact quotation was not searched for in the reports, but a description in the CCIPS report of a person taking the action described, such as a hacker obtaining login credentials through a phishing email, would qualify the variable to be marked as "yes."

The "built/maintained software" variable was marked as "yes" if the article indicated that a program that the attacker built, or some technical component that the attacker maintained (such as a website or botnet) was used for the hack. This variable was also marked as "yes" if an individual used someone else's code, but significantly modified it. For example, if a hacker found malware on the dark web and then changed it to suit his/her purposes, the "built/maintained software" variable was still a "yes." The variable of follow-up access was identified in the report if the article detailed any hacks that could only be completed because of

the success of a previous hack committed by the same perpetrator. This does not include multiple hacks that do not build off of each other, so if a perpetrator hacked two separate accounts, even if the same method was used for both hacks, this would not be considered follow-up access. It was only counted as follow-up access if the second hack was accomplished using information or access from the first hack.

The "targeted multiple entities" variable was included if the article made note of multiple entities that the attacker targeted. Similarly, if an entity that the attacker targeted was an organization, then the "targeted an organization" variable was marked as "yes." This study determined that an organization could be a company, government branch, or any formal group of people. If the article gave a clear indication that there was a political or nationalistic motive for the hack then the political/national element variable was marked as "yes."

In each entry, the "number of participants" variable was marked as "individual" if the article mentioned only one perpetrator, "pair" if it mentioned two perpetrators, and "group" if it mentioned three or more perpetrators. Although this variable was not used for analysis, a variable recording whether or not the attacker worked alone, which was extrapolated from the "number of participants" variable, was recorded. Finally, as oftentimes a single CCIPS report contained information about multiple perpetrators, this paper also recorded whether or not a single article was used to code multiple individuals in the "repeat articles" variable. Every time an article was used for the first time this variable was marked as "no" but for each additional time the same report was used the variable was marked as "yes." It was not used for any analysis but is listed in the descriptive statistics table.

A potential issue with this paper's approach is the fact that, given the nature of the USDOJ press releases, there was not always enough information to determine whether a variable applied to a given case. In some cases where a variable was not explicitly stated, but the situation that the report described gave a clear answer to the variable, that observation was imputed. However, with the exception of the "number of perpetrators," "age," "international resident," and "repeat article" variables, if a variable was not explicitly given by the report, then it was marked as "no" in the final analysis. If a report described multiple hacks attributed to the same individual, then a variable was marked as "yes" if any of the described hacks contained indicators of a given variable.

Another variable, "attack sophistication" was created by combining the "insider information," "social engineering," "built/maintained software," "follow-up access," "targeted multiple entities," and "targeted an organization" variables. The variable was an original creation for the purposes of examining correlations between hacker characteristics and technical ability using the variables identified in this study. To calculate attack sophistication, each cyberattack was checked for whether social engineering was used in a hack, if the hack targeted an organization/multiple entities, and if the attack involved building/maintaining software or follow-up access. Each of the above variables were summed together into a scale such that if the characteristic was present, one point was added to the attack sophistication score. An additional variable, whether the attack used insider information was also included, but answering "yes" for this variable subtracted one point from the attack sophistication score, because an attack completed with insider information is easier for the hacker than an attack without insider information. A higher score indicates a more sophisticated attack, and a lower score indicates a less sophisticated attack.

Although none of these variables individually are a clear indicator of how advanced a hack was, a rough estimate of the technical sophistication of the hack can be ascertained when all these variables are used together. For example, a hack that was completed with insider information can still be a more advanced hack than a separate hack that was completed without insider information. However, assuming the only difference between two hacks is that one was accomplished with the help of insider information, and one hack was not, then the hack accomplished without insider information would be considered more sophisticated.

It is this idea, that multiple variables can be used together to implicate hack sophistication, that informed the final analysis of the paper. Obviously, different variables might not be the same indicator of technical proficiency as other variables (e.g., hacks performed where attackers built software might be, on average, more sophisticated hacks than attacks with multiple targets), yet in terms of the "attack sophistication" variable, all other variables were treated as equal indicators of hack proficiency. Hacks that were completed by multiple perpetrators posed an issue to this analysis, because groups or pairs would likely have greater total technical proficiency than individuals. To address this issue, whether or not a hacker worked alone was included as an independent variable in the regression test for the "attack sophistication" variable.

Analysis of variables checking if a hack had a political/national element or if the perpetrator of a hack worked alone was also conducted for profiling purposes. To this end, these variables were tested against the variables of age, gender, and residency to determine any correlations. If any significant correlations were present, they were recorded as potential ways that details of a hack could predict the age, gender, or residency of the hacker. The six variables used to determine hack sophistication were also used in profiling analysis so that if a certain element of a cyberattack is known (for example, that the perpetrator used social engineering in their attack), law enforcement would be able to predict the likelihood that the hacker is older, a woman, or an international resident. Finally, regression tests were performed to see how each of the eight profiling variables and the other characteristic variables independently affected age, gender, and residency of a hacker.

Results

This study analyzed a total of 54 different hacking related cases involving 122 individuals. Out of these cases eight individuals were identified as female and 95 were identified as male. For 19 individuals, gender was not specified in the CCIPS report. 36 hacks were completed by an individual working alone, while the rest were completed by groups (n=71) or pairs (n=15). Exactly half of the 122 hacking perpetrators were identified as international residents, and only two cases existed of residency of a hacker not being identified. These results, as well as descriptive statistics for hacking characteristics present in the reports, can be seen in Table 1.

Age of alleged hackers ranged from a minimum of 19 years old (juveniles were excluded in the data set) to a maximum age of 66. The largest age range group was the 28-37 range, which included 48 out of the 108 participants with a recordable age. The smallest age range was the 48-57 group, which included just three individuals (Table 2).

Table 1. *Descriptive Statistics for Nominal Variables Valid (N =122)*

	N	%
Number of Participants		
Individual	36	29.5
Pair	15	12.3
Group	71	58.2
Repeat Articles		
No	77	44.3
Yes	45	55.7
Gender		
Male	95	92.2
Female	8	7.8
International Resident		
No	59	48.4
Yes	61	50
Political/National Element		
No	110	90.2
Yes	12	9.8
Insider Information Used		
No	97	79.5
Yes	25	20.5
Social Engineering		
No	78	63.9
Yes	44	36.1
Built/Maintained Software		
No	81	66.4
Yes	41	33.6
Follow-Up Exploit		
No	69	56.6
Yes	53	43.4
Multiple Entities Targeted		
No	40	32.8
Yes	82	67.2
Organization Targeted		
No	31	25.4
Yes	91	74.6

This study found many relationships between characteristics of an attack and age of an attacker. Using t tests, it was found that, on average, a hacker who used social engineering is four years younger than a hacker who did not. A hacker who built or maintained software in his/her attack is about four years older than a hacker who did not, and hackers using follow-up access in a hack are about four years younger than their older counterparts. When using linear regression to determine the relationship between characteristics of an attacker/attack and age, many relationships were discovered. Follow-up access in an attack lowered the age of an attacker by about five and a half years, and building/maintaining software increased the age of

an attacker by about ten years (Table 3). This means that younger hackers use follow-up access more often, and older hackers tend to build and maintain software in their attacks. Other factors, such as a hack targeting an organization/multiple entities, using insider information, using social engineering, whether the hacker worked as an individual, the gender of a hacker, and the residency of a hacker were not found to be significant in predicting the age of the attacker.

Table 2. *Descriptive Statistics for Age (N=108)*

	%	N	
Age Ranges			
18-27	29.6	32	
28-37	44.4	48	
38-47	19.4	21	
48-57	2.78	3	
58+	3.7	4	
	Mean	SD	Range
Age	33.19	9.631	47

Furthermore, this study found relationships between characteristics of an attack and whether an attacker was an international resident. Using chi-square tests, it was found that, on average, a hacker who committed a hack involving a political/national element, social engineering, targeting of multiple entities, or building/maintaining software was more likely to be an international resident than a domestic one.

Table 3. *Logistic Regression for Determinants of Residence*

	Model # 1
Independent Variables	
Social Engineering	-1.474
Built/Maintained Software	9.583***
Follow-Up Access	-5.508*
Targeted Multiple Entities	-.835
Targeted Organization	2.735
Insider Information	-.807
Worked Alone	2.910
Political/National Element	-7.323
Gender	-5.930
International Resident	-2.267
Constant	38.004
r2	.259
N	99

Note: *** p<.001, ** p<.01, * p<.05

Table 4. *Residency of a Hacker by Use of Insider Information*

	Residency of Hacker	
Insider Information	Frequency of Internationals	Frequency of Nationals
Insider Information not Used	60	35
Insider Information Used	1	24

Note: $\chi 2$ (1) = 27.713, p = 0.0000001407

Conversely, it was found that hackers who committed hacks with insider information and hackers who worked alone were less likely to be international residents (Table 4 and Table 5). Specifically, approximately 47 percent of the total domestic attacks (calculated from numbers in the rightmost column of Table 5) were conducted by individuals, while the remainder were conducted by pairs/groups.

Table 5. *Residency of a Hacker by Use of Working Alone*

	Residency of Hacker	
Worked Alone	Frequency of Internationals	Frequency of Nationals
Did Not Work Alone	55	31
Worked Alone	6	28

Note: $\chi 2$ (1) = 20.905, p = 0.000004826

There were 61 internationally based attacks, of which 56, or 90 percent, conducted the attack within a group. Chi-square analysis confirmed the association as highly significant. When using binary logistic regression to predict residency, it was found that attackers who built or maintained their own software were over five times more likely to be international (p< .018), but other relationships found previously with chi-square analysis were no longer significant when logistic regression was applied (Table 6).

Table 6. *Logistic Regression for Determinants of Residence*

Social Engineering	.807
Built/Maintained Software	5.510*
Follow-Up Access	.444
Targeted Multiple Entities	1.961
Targeted Organization	1.810
Insider Information	.123
Worked Alone	.308
Political/National Element	666266800.68
Gender	491500000000000
Age	.972
Constant	.904
r2	.526
N	99

Note: *** p<.001, ** p<.01, * p<.05

There also were relationships between the gender of a hacker and characteristics of an attack. Using chi-square tests, it was found that if an attacker was an international resident or targeted multiple entities in their hack, then that individual was less likely to be female (see Table 7 and Table 8).

Table 7. *Gender of a Hacker by Residency of a Hacker*

	Gender of a Hacker	
Residency of Hacker	Frequency of Males	Frequency of Females
National Resident	47	7
International Resident	47	0

Note: $\chi 2$ (1) = 6.546, p = 0.01051

When using binary logistic regression to check variables for an effect on gender, no relationships were found. However, it is worth noting that in this study, results between gender and characteristics of an attacker were much harder to find, because there were only eight females in the sample, as opposed to 95 men. This implies that men are much more likely to be perpetrators of a cyberattack than women are.

Table 8. *Gender of a Hacker by Targeted Multiple Entities in Hack*

	Gender of a Hacker	
Targeted Multiple Entities	Frequency of Males	Frequency of Females
Did not Target Multiple Entities	29	6
Targeted Multiple Entities	47	0

Note: $\chi 2$ (1) = 6.505, p = 0.1076

Finally, tests were conducted using linear regression to determine how the age, gender, and residency of an attacker, and whether said hacker worked alone, predicted the attack sophistication. Through these tests it was found that international hackers tend to have more sophisticated cyberattacks, but there is an inverse relationship between whether a hacker works alone and the sophistication of the cyberattack that they commit (see Table 9).

Table 9. *Linear Regression for Determinants of Attack Sophistication*

Worked Alone	-1.360***
Age	.001
Gender	.544
International Resident	.862**
Constant	1.775
r2	.365
N	99

Note: *** p<.001, ** p<.01, * p<.05

Discussion and Conclusion

This study sought to bridge the divide between technical and social science within cybercrime, by studying the connection between a hack and personal characteristics of the hacker. The main purpose in doing so was to aid digital forensics investigators by providing profiling information. A secondary purpose was to

discover causal relationships between aspects of a hacker and methodology of a hack. To accomplish this goal, 122 USDOJ CCIPS reports were reviewed and analyzed, an experiment that provided useful findings.

Perhaps the most unique results to come out of this analysis was the understanding of how age, gender, residency, and working alone affect attack sophistication. Age and gender were found to have no significant effect on attack sophistication. These findings might reinforce an earlier study's conclusion that age and gender do not affect success rate of a hack (Bachmann, 2010), since if calculated correctly, it is feasible that attack sophistication would be related to attack success.

Individuals working alone were found to have less sophisticated attacks than individuals working in pairs or groups, echoing a prediction by Chng et al. (2022) that attack sophistication would have a connection with whether or not a hacker worked alone. This finding is logical, since more people working on a hack would naturally allow for both more ideas and more resources to be put into attack planning and execution. Additionally, international residency was found to have a positive relationship with attack sophistication, a correlation that is elaborated on later in this section.

Although there are already known correlations between characteristics of hackers and their attacks (e.g., Bachmann, 2010; Back et al., 2019; Woo, 2003), such as the fact that hackers who think analytically are more successful than hackers who do not (Bachmann, 2010), many of those relationships found in this study disappear when using a test which accounts for the presence of other variables. In other words, some characteristics of an attack might correlate with characteristics of an attacker, but there is no causal relationship between the characteristics of the attacker and the attack. This can be seen with tests concerning residency of an international hacker.

There was positive correlation between a hacker being an international resident and the attack he or she committed involving social engineering, but in a binary regression test this relationship was no longer significant. This was also true for most of the relationships between hack characteristics and residency found using chi-square tests. The fact that this was true for social engineering implies that there is nothing about a hacker being an international resident that causes them to commit more hacks using social engineering, rather because of other factors international residents end up using social engineering in their attacks more. In order to determine these other factors, chi-square tests were conducted to see how social engineering related to other variables in the study. These tests found that use of social engineering correlates with a higher chance of building/maintaining software. This is the same relationship that building/maintaining software has with residency. Therefore, the "relationship" between use of social engineering and residency could just be a result of the fact that there is a relationship between building/maintaining software and social engineering. It is also possible that many of the relationships found with chi-square tests, including social engineering, become no longer significant because the sample size is not large enough for the relationship to hold with the addition of more variables. Regardless, more research should be conducted in this area to determine the correct answer and discover which correlations, if any, have a causal effect.

However, there is still a relationship that exists between residency and characteristics of a cyberattack after accounting for other variables such as gender, age, if an attack had a political element, targeted an organization, targeted multiple entities, used follow-up access, used insider information, used social engineering, used built/maintained software, and if the attacker worked alone. The main finding after accounting for

these variables is that international attackers are more likely to build or maintain software in their attacks, a fact that can potentially be used by digital investigators to ascertain the likelihood that perpetrators of certain hacks are international residents. It is unclear why exactly this relationship exists, but one explanation that this paper proposes is that the USDOJ data being used presents bias. The data used for this report comes from USDOJ cases, so these cases exclusively affected United States individuals and organizations. It would most likely be easier for hackers to target an entity within their country of origin than in the United States, as a hacker would supposedly have more knowledge about people or organizations within their homeland. It may be the case that an international resident targeting an entity in the United States is already selecting a harder target and would need greater resources (such as specialized built/mantained software) to successfully commit such an attack. If the explanation proposed is accurate, it would also explain why, as mentioned before, international hackers were more likely to have more sophisticated cyberattacks – because it may be the case that attacking U.S. targets requires it. Furthermore, although use of insider information is not significant in a regression test (however it is close to being significant, with a p value of .091) it is significant in a chi-square test with residency. The results imply that international residents attacking U.S. entities would be less likely to have insider information about said entity. This could potentially be the result of international residents having less access to U.S entities than nationals, and consequently being less likely to acquire insider information. If true, that rationale would provide further evidence for the suggestion that international residents need more resources to attack U.S. entities, in order to compensate for the lack of information.

Another possible explanation for the phenomenon of international hackers having more sophisticated attacks than nationals is that there exists some correlation between nationalism and hacking skill. Prior research has shown that nationalistic hackers are more likely to hack enemy nations (Woo, 2003), so it may be the case that international hackers from nations antagonistic to the United States (a large portion of the data set) are more nationalistic than the average hacker. A potential reason for such a correlation could be that nationalistic hackers might be more encouraged to develop their hacking skills than other hackers, or simply that nationalistic hackers attacking enemy countries may possess increased motivation during their hacks. Further research is necessary to test either hypothesis for the apparent skill difference between national and international hackers.

There are also some relationships present when viewing the effects of linear regression on the age of an attacker, namely that even when accounting for other variables, follow-up access and building/maintaining software have a significant effect on the age of an attacker. The fact that there is a relationship between age and building or maintaining software should not be surprising – building or maintaining software is something that takes experience in software development, and research has shown that programmers become more skilled as they advance in their careers (Morrison & Murphy-Hill, 2013). Older people would have more time to advance in programing careers and, according to this research, might therefore be more skilled programmers. As a result, they may be more likely to use their software development skills more often in hacks. The fact that there is also a relationship between committing an attack with follow-up access and age might appear to be less intuitive, but there are two explanations for this relationship that this paper proposes.

First, it is possible that attackers who commit attacks with follow-up access (previously defined in this paper as "a hack that could only be completed because of the success of a previous hack committed by the same perpetrator") are more aggressive, because they are making their initial hacks larger. Aggressive hackers might be more likely to be caught by law enforcement, and sent to prison, thereby making it less likely that they will continue to hack. This idea could be supported by prior research indicating that hackers with a low risk propensity (who are more cautious and might, as a result, be less aggressive) are more successful than those with a high risk propensity (Bachmann, 2010), as it is possible that more successful hackers are better at evading prosecution. If this suggestion is accurate, the relationship between follow-up access and age might be the result of a "funnel" effect, whereby aggressive hackers willing to take risks are forced to stop hacking, and the hackers that do not use follow-up access are more likely to evade detection and continue hacking when they are older. However, an issue with this explanation is that if it is true, there should also be a relationship between age and targeting multiple entities, since targeting multiple entities is behavior that can also conceivably be seen as risky and aggressive. There was no such relationship found, which presents a challenge to that idea.

A second possibility for the connection between age and follow-up access is that older hackers, who are potentially more mature and experienced, may be more likely to be deliberate in their attacks. Follow-up hacks might be something that some hackers do when they see the opportunity arise, while more disciplined hackers remain focused on their initial goal. It could be that older hackers are more disciplined, and therefore less likely to "stray from the path" to add additional hacks. However, this explanation has the same issue described previously, because if older hackers are less likely to have "additional" hacks, they should also be less likely to target multiple entities. Regardless, this is an area where future research should be conducted.

A final point of interest in the data is the fact that generally no relationship exists between gender and characteristics of a cyberattack. This is probably because there was a very small number of female hackers in the data set (only eight out of 122). Consistent with prior research (Bachmann, 2010; Steinmetz, 2015; Turgeman-Goldschmidt, 2008), these findings imply that hackers are more likely to be male, although it is also possible that female hackers are better at evading detection or that the USDOJ handles female hackers differently. It is also unclear from the reports how gender was identified. Whether the subjects of the reports were prompted to give their gender identity, and if so, what options were provided to the subjects, was not information supplied in the data set.

In addition to the limitations on gender discussed, limitations of this study mainly came from the fact that the USDOJ reports used did not always include all necessary information about relevant cyberattacks. As a result, the variables recording whether attackers used built/maintained software, social engineering, insider information, and/or follow-up access in their hacks were often unclear in the reports. For example, what counted as "building or maintaining software" would depend on how "maintaining" software was defined. An individual using malware found on the dark web (and potentially tailoring that software to suit the hacker's needs) could reasonably count as either "yes" or "no" for that variable. Additionally, the "insider information" variable was uniquely difficult to code, since insider information could be obtained by a hacker through a variety of techniques that would not necessarily be stated in a USDOJ report.

Furthermore, this paper defined "hacks" specifically as unauthorized access to simplify results, but it was not always clear from the data what constituted unauthorized access. In particular, some reports stated that an individual had committed a cyberattack, but it was unclear whether or not the attacker needed "unauthorized access" for the attack, so those reports were not included in the data set. Many of these decisions were subjective and based on the researcher's understanding of the reports, so a similar study conducted by others might have different results. Likewise, the metric used to create the "attack sophistication" variable was also subjective, and a different measurement for attack sophistication could affect the analysis.

Further issues with the data involved the presence of multiple hackers. Many hacks reviewed were completed by groups or pairs, and it was difficult to determine which members of the group committed which parts of an attack. If the report did not specify individual roles in the hack, then all members of the group were counted as having done all parts of an attack. However, if a report specified what individual members did, then each member of the group was coded differently. This might have led to some issues in the accuracy of data. It should also be noted that if details of an attack were unclear, or even if there was a preponderance of evidence that the attack had a certain characteristic, certain variables were marked as not being present in the attack, which could have led to false negatives.

Another limitation is that age was given at the time of the report, not at the time of the attack. Due to the fact that the U.S. punitive system takes a long time to work, the ages of the individuals in this report were older by one to five years because of different phases being experienced within the criminal justice system (indicted, charged, arraigned, plead guilty, etc.). A more consistent aging system could have changed results involving the age variable.

A final limitation of this study is that it depended on what the USDOJ chose to publish, which were only individuals who were apprehended and tried in the criminal justice system, so many criminals and crimes were not included. As cybercrime is notoriously difficult to detect (Lee & Holt, 2020) and is often underreported (Amir et al., 2018), similar limitations are an issue for any cybercrime data set. Furthermore, the decision by the USDOJ to publish information on certain cases and individuals might have also influenced the data. More research on the relationships between cyberattacks and characteristics of attackers can be conducted in the future with different data sets in order to eliminate this issue.

Despite the limitations of this study, certain discoveries were made. The results imply that hacks completed by international citizens possess certain indicators that perpetrators of the cybercrime are not U.S. citizens, which might be useful in cybercrime investigations. It was also found that there is a relationship between older hackers, and building or maintaining software in their hacks, which can be used to understand how cybercrime type and complexity varies across different groups. Finally, although there were not many correlations found for female hackers, certain correlations were found that could be useful in profiling, and give hints as to where the differences between male and female hackers lie. Hopefully, as the field of cybercrime continues to grow, more information can be found about hackers, and society can better defend itself against this dangerous new type of crime.

Acknowledgements

Gratitude is expressed towards University of Florida professor Dr. Tiffany Jenson (this project's primary research advisor), as well as professors Dr. Joseph Wilson, Dr. Joseph Rivera, and fellow researcher Lillian Phillips for their input on this project.

References

Amir, E., Levi, S., & Livne, T. (2018). DO firms underreport information on cyber-attacks? evidence from Capital markets. *SSRN* Electronic Journal.

Bachmann, M. (2010). The risk propensity and rationality of computer hackers. *International Journal of Cyber Criminology, 4*, 643-656.

Bachman, R., & Schutt, R. K. (2008). The processes and problems of criminological research. In *Fundamentals of research in Criminology and Criminal Justice* (pp. 27–56). essay, SAGE.

Back, S., LaPrade, J., Shehadeh, L., & Kim, M. (2019). Youth hackers and adult hackers in South Korea: An application of cybercriminal profiling. *2019 IEEE European Symposium on Security and Privacy Work shops (EuroS&PW)*, pp.410-413, Stockholm, Sweden.

Back, S., Soor, S., & LaPrade, J. (2018). Juvenile hackers: An empirical test of self-control theory and social bonding theory. *The International Journal of Cybersecurity Intelligence and Cybercrime, 1*(1), 40–55.

Charlesworth, T. E. S., & Banaji, M. R. (2019). Gender in science, technology, engineering, and mathematics: Issues, causes, solutions. *The Journal of Neuroscience, 39*(37), 7228–7243.

Chiesa, R., Ducci, S., & Ciappi, S. (2018). To Be, Think, And Live As A Hacker. In *Profiling hackers* (pp. 33–56). Auerbach Publications.

Chng, S., Lu, H. Y., Kumar, A., & Yau, D. (2022). Hacker types, motivations and strategies: A comprehensive framework. *Computers in Human Behavior Reports, 5*, 100167.

Cliffe, D. (2018). Thriller, horror, hacker, spy: *The 'hacker' genre in film and television from the 1970s to the 2010s*. Thesis.

Donner, C. M. (2016). The gender gap and cybercrime: An examination of college students' online offending. *Victims & Offenders, 11*(4), 556–577.

Greenwood, S. (2022, December 15). *Internet, smartphone and social media use*. Pew Research Center's Global Attitudes Project. Retrieved March 5, 2023, from https://www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/

Grispos, G. (2019). Criminals: Cybercriminals. *Encyclopedia of Security and Emergency Management*, 1–7.

Hadzhidimova, L. I., & Payne, B. K. (2019). The profile of the International Cyber Offender in the U.S. *The International Journal of Cybersecurity Intelligence and Cybercrime, 2*(1), 40–55.

Holt, T. J., & Steinmetz, K. F. (2020). Examining the role of power-control theory and self-control to account for computer hacking. *Crime & Delinquency, 67*(10), 1491–1512.

Hutchings, A., & Holt, T. J. (2018). Interviewing cybercrime offenders. J*ournal of Qualitative Criminal Justice & Criminology, 7*(1).

Jaquet-Chiffelle, D.-O., & Loi, M. (2020). Ethical and unethical hacking. *The International Library of Ethics, Law and Technology, 21*, 179–204.

Lee, J. R., & Holt, T. J. (2020). Assessing the factors associated with the detection of juvenile hacking behaviors. *Frontiers in Psychology, 11*.

Miró-Llinares, F., & Moneva, A. (2019). What about cyberspace (and cybercrime alongside it)? A reply to Farrell and Birks "did cybercrime cause the crime drop?" *Crime Science, 8*(1), 1-5.

Morrison, P., & Murphy-Hill, E. (2013). Is programming knowledge related to age? an exploration of stack overflow. *2013 10th Working Conference on Mining Software Repositories (MSR).*

Payne, B. K., & Hadzhidimova, L. (2020). Disciplinary and interdisciplinary trends in cybercrime research: An examination. *International Journal of Cyber Criminology, 14*(1), 81-105.

Payne, B. K., Hawkins, B., & Xin, C. (2018). Using labeling theory as a guide to examine the patterns, characteristics, and sanctions given to cybercrimes. *American Journal of Criminal Justice, 44*(2), 230–247.

Shoemaker, D., & Kennedy, D. B. (2009). Criminal profiling and cyber criminal investigations. In F. Schmalleger & M. Pittaro (Eds.), *Crimes of the internet* (pp. 456–476). essay, Pearson, Prentice Hall.

Steinmetz, K. F. (2015). Becoming a hacker: Demographic characteristics and developmental factors. *Journal of Qualitative Criminal Justice & Criminology, 3*(1), 31-60.

Summers, T. C. (2015). *How hackers think: A mixed method study of mental models and cognitive patterns of high-tech wizards.* Thesis.

Turgeman-Goldschmidt, O. (2005). Hackers' accounts. *Social Science Computer Review, 23*(1), 8–23.

Turgeman-Goldschmidt, O. (2008). Meanings that hackers assign to their being a hacker. *International Journal of Cyber Criminology, 2*, 382-396.

Woo, H.-J. (2003). *The hacker mentality: Exploring the relationship between psychological variables and hacking Activities*. Thesis.