Datasets for Analysis of Cybercrime

10

C. Jordan Howell and George W. Burruss

Contents

Abstract

In this chapter, we document various sources of cybercrime data to help guide future research endeavors. We focus most of our attention on datasets associated with hacking, and to a lesser degree online fraud. Rather than a catalog of sources, we also describe what research has accomplished with these data on specific crimes and discuss the strengths and limitations of their use. The data discussed throughout the chapter are gathered from a variety of sources including the FBI, Cambridge Cybercrime Centre, Zone-H, various cybersecurity companies, and several other websites and platforms. These data allow researchers the opportunity to assess cybercrime correlates of engagement, victimization patterns, and macro-level trends. However, they share one major flaw; they do not allow for the assessment of causation. We conclude by suggesting that criminologists should prioritize longitudinal data collection that allows for causal assessment.

Keywords

Cybercrime · Datasets · Analysis

C. J. Howell · G. W. Burruss (✉)
Department of Criminology, University of South Florida, Tampa, FL, USA
e-mail: cjhowell@mail.usf.edu; gburruss@usf.edu

T. J. Holt, A. M. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, https://doi.org/10.1007/978-3-319-78440-3_15

Introduction

Criminologists use official data to examine trends and correlates of criminal behavior, especially in developed countries where such data are routinely and systematically collected. The US Uniform Crime Reports (UCR) is an exemplar of official data collected at the local level and aggregated at the national level via the Federal Bureau of Investigation (FBI 2019a). Using this kind of data, criminologists examine the changes in crime rates over time and across jurisdictions for both property and violent crimes, correlating change with structural factors like unemployment, demographics, and policing practices. Analysis of these data also allows testing of criminological theories, such as routine activities, institutional anomie, and general strain. However, official data suffer from the "dark figure" of crime where much goes unreported by victims, unfounded by the police, or misclassified by the reporting agency. It should be noted that in the United States, the National Incident-Based Reporting System (NIBRS) has been developed to address many of the shortcomings of the UCR system (see FBI 2019b). These kinds of official datasets represent secondary data analysis where records collected for one agency or researcher's purpose can be used to answer myriad research questions not considered by the collecting agency.

To uncover the dark figure of crime, criminologists also use officially collected data that do not rely on official victim reporting to the police. Annual victimization surveys, like the Crime Survey for England and Wales (see Office for National Statistics 2019), attempt to estimate crime through a nationally representative household survey (approximately 50,000). These data constitute self-reported criminal experiences by victims. In addition to capturing crime unreported to the police, victimization surveys provide individual information about victims that criminologists can use to find correlates and causes of crime. The data can also be used to determine the level of underreporting in official police data. Victimization data are also considered secondary given the data are collected systematically for one agency's purposes. Individual researchers, however, often conduct their own victimization surveys, which would constitute primary data collection (e.g., Schafer et al. 2018).

For cybercrime, the same issues with official and victimization data for traditional crimes are applicable, but cybercrimes have more in common with white-collar crime than traditional crimes. For traditional crime, the more serious or harmful the offense, the more likely it will be reported to police. For example, most homicides are recorded in the UCR, and much information about the criminal events (victims, offenders, mode of death, and location) are available in the UCR's Supplemental Homicide Reports. About 60% of US homicides are cleared by arrest or exceptional means (FBI 2019a) meaning we know much about the details of the offense. For less serious crime, like larceny-theft, only about 19% are cleared by arrest or exceptional means; thus, we know much less about the true amount of this crime that occurs year to year. For white-collar crime, many victims fail to report the crime to the police unaware they have been swindled, or they are ashamed, having willingly given away money or something of value through deception. White-collar

crimes are not reported in the UCR Part 1 offenses, and only a few are reported in the Part 2 offenses. The same reporting issues face general victimization surveys as respondents are often unaware of their own victimization, and most general victimization surveys only capture a few fraud types (but see Huff et al. 2010 for an example of a white-collar specific victimization survey).

Cybercrime victims are similar to fraud victims because they often fail to realize they have been wronged. Even self-reported victimization surveys may fail to record incidents of cybercrime given the victims are often naïve about computer-related offenses. For example, most victims of a business' data breach are only made aware through the media's reporting or because they are contacted long after the fact by the business. A victimization survey that asked specifically if a hacker had access to personal account information, passwords, or banking details would yield many false negatives until the breach was made public. Even then, the victims may not be sure they were in fact victimized. Nevertheless, cybercrime is unlike most white-collar or traditional crime in that networked computers generate vast amounts of data on user usage, social networks, and online behavior, such as shopping, dating, and posting information. As criminologists study cybercrime in more depth, they are just now beginning to use this array of data sources to understand the motives of offenders, lifestyles of victims, and the interaction of computing devices and human behavior.

In this chapter we focus on available datasets, divided into three sections: official, proprietary, and open-source data. Though we cover various sources of information in this chapter, our coverage is not exhaustive of the data available at the time of this writing. Also, our coverage is not comprehensive regarding various types of cybercrimes given the space limitations. For example, we do not address cyberbullying or cyberstalking and the potential sources of data for these offenses available to researchers. As a result, we focused most of our discussion on datasets associated with hacking and to a lesser degree online fraud.

Official Data Sources

Official data sources, for the purpose of the current chapter, include data published by government agencies or other organizations working on behalf of a government agency. Official data sources can be further subdivided into police data and victimization data. As previously mentioned, the UCR is the largest source of official police data in the United States. Local law enforcement agencies across the United States report crimes that occurred within their jurisdiction to the FBI. One major strength of the UCR data source is that it includes crime data for the majority of agencies within the United States. Researchers can use these data to gain insight into 21 different crime types. Unfortunately for cybercrime scholars, the UCR does not report cybercrime incidents. Therefore, the UCR cannot be used as a reliable source of official cybercrime data.

Fortunately, NIBRS was developed to improve upon the UCR. Similar to the UCR, NIBRS reports crimes that are known to the police. However, NIBRS includes more crime types and more information about the crime. Specifically, NIBRS gathers data on victims, known offenders, arrestees, and the relationship between offenders and victims. NIBRS can be used to study cybercrime in two ways. First, researchers can use NIBRS data to determine if a computer was used during the commission of a crime. Second, NIBRS reports detailed information for reported online fraud incidents including hacking and wire fraud. Therefore, researchers can use these data to gain insight into online fraud cases that have been reported to the police. Although NIBRS provides detailed information for a wide range of crimes, a large majority of cybercrimes are not captured by NIBRS. In addition, as stated above, NIBRS only captures known offenses. Therefore, NIBRS underestimates crime generally and cybercrime specifically.

The most prominent repository for official counts of cybercrime incidents in the United States is the Federal Bureau of Investigation's *Internet Crime Complaint Center* (IC3). The IC3 was started in 2000 to provide the American public a convenient way to report cyber victimization. Since its inception, over 1,400,000 complaints have been filed, totaling over $5.5 billion in total documented losses (FBI 2017). However, due to underreporting by victims, this is likely only a fraction of the total number of incidents that have truly occurred (Button et al. 2014).

Initially designed as a tool for law enforcement, the IC3 now allows researchers to download its data. Researchers can examine city, state, county, and country reports and sort the data by crime type, age, and transactional information. While useful for understanding emerging trends (Pangaria and Shrivastava 2013), the data are limited in the capacity to correlate potential causal factors, such as structural, demographic, or policy changes. Nevertheless, as a secondary dataset, the IC3 offers potential explanatory power when combining cybercrime trends with other sources of official data, such as from the US Census Bureau. As noted above, the problem with reporting bias makes the IC3 data problematic, but certainly an important step in understanding the distribution of American cybercrime victimization.

In addition to analyzing cases reported to the police, researchers can simply ask individuals if they have been victimized. The most well-known victimization survey in the United States is the National Crime Victimization Survey (NCVS), which is administered by the Bureau of Justice Statistics. Although not designed to study cybercrime specifically, the NCVS includes questions regarding online identity theft. In addition, the NCVS collects information about cyberbullying as part of its School Crime Supplement. Data from the NCVS can be used to garner insight into the prevalence of victimization, how the crime is discovered by the victim, the financial and nonfinancial burden caused by victimization, whether the crime was reported, and if actions were taken to reduce future incidents. Although the NCVS has been successfully used to explain American victimization patterns and test criminological theory (Muniz 2019), it is limited in its ability to explain cyber victimization patters due to the limited amount of information gathered about a limited number of cybercrimes.

In addition to the NCVS, the National Computer Security Survey (NCSS), which is cosponsored by the Bureau of Justice Statistics and the National Cyber Security Division (NCSD) of the US Department of Homeland Security, can be used to gain insight into computer security incidents. The NCSS was designed to gather and produce reliable estimates of computer security incidents against American businesses. In 2005, the NCSS gathered data for 7818 businesses. The data can be used to estimate monetary losses, system downtime resulting from cyber incidents, the types of offenders, whether the incident was reported, the types of systems infected, and the most commonly exploited vulnerabilities (Rantala 2008). Unfortunately, the most recent data was gathered in 2005, which limits the data source's utility in understanding cybersecurity incidents in contemporary times.

Although the aforementioned data sources provide information about cybercrime in the United States, multiple other countries collect similar data. For example, Canada's General Social Survey (GSS) was developed in 1985 to monitor the well-being of Canadians and guide policy. The GSS asks those selected to participate in the survey to answer a wide array of questions about victimization generally and various forms of cyber victimization more specifically. The data can be used to analyze victimization patterns for both cyber-dependent crimes (i.e., hacking, phishing, malware infection) (Reyns 2015) and cyber-enabled crimes (i.e., cyberbullying, cyberstalking) (Reyns et al. 2016). In addition, the Crime Survey for England and Wales asks people living in England and Wales about their experiences with crime in the past 12 months. In regard to cybercrime, the survey includes questions about online fraud victimization. Using these data sources, and similar data sources collected in other parts of the world, researchers can analyze the longitudinal trends of cybercrime victimization. Although the reliability of the data may vary nation to nation, these datasets can be analyzed independently or combined to advance our understanding of global trends.

Proprietary Data

In addition to official criminal justice data, businesses, for-profit data collection enterprises, and nongovernmental agencies also collect information about cybercrimes and victimizations. Cybersecurity has become a major global industry – from selling antivirus software to consulting on risk management. These businesses collect an immense volume of data to support their cybersecurity ventures. Companies like McAfee, Norton, and Kaspersky employ antispam and virus protection software across millions of business and personal computing devices. This software in turn collects information about hacking attempts, DDoS attacks, malware installation, spam campaigns, and phishing attempts. Because much of this information is proprietary, these companies do not typically release the data to researchers, though there are notable exceptions explained below.

There are also ventures by enterprising individuals to collect open-source and self-report data that can be used for research. Though often problematic because the

data collection was not devised by careful researchers, these kinds of data sources can prove to be useful. Finally, nongovernmental organizations, such as research consortiums, can engage in data collection across many sources to create data archives open to either the public or other researchers. In this section we discuss some of these kinds of proprietary data sources.

While much proprietary data generated by cybersecurity businesses are not made public, some companies do make aggregated data available for public consumption. An example of a publicly available threat map is that produced by *Kaspersky Lab*, a multinational cybersecurity company and antivirus provider. As of 2016, their software had over 400 million users. The company constantly scans for cybersecurity-related incidents and displays them in "real time" (Kaspersky Lab 2018). With over 400 million users, data gathered from Kaspersky Labs can be used to show country-level variation in victimization patterns and overall global trends. Currently, Kaspersky Lab collects data on the frequency of local infections, web threats, network attacks, vulnerabilities, spam, infected mail, and botnet activity.

Like Kaspersky Lab, *Bitdefender* is a cybersecurity company that offers data visualization in the form of an interactive map. With over 500 million users, this is another invaluable source of attack and victimization data. Bitdefender (2018) collects data on the total number of infections, attacks, and spam at the country level. Viewers are able to see where the attack originated and the target country. However, researchers should be cautious when examining the origin of a cyberattack from any data source because hackers often hide their location through looping, using one computer to access another (Lee et al. 1999).

McAfee is an American-owned cybersecurity software company. McAfee offers an interactive map similar to those discussed above. The map provides country-level attack and spam data. In addition, McAfee provides malware data that can be analyzed at the macro-level and the event level. Specifically, they show how much malware each country receives and offer event-specific information such as the type of malware, the associated risk, and the date and time of discovery (McAfee 2018).

In addition to companies that generate threat maps, other private cybersecurity businesses provide data on malicious online activity. *Trend Micro* is a multinational cybersecurity company that continuously monitors network activity to identify command-and-control (C&C) servers. A C&C server is a centralized computer that controls botnets, which are Internet-connected devices done so without the owners' permission. These C&C servers can be used to create an army of infected computers often used in data breaches, DDoS attacks, and a number of other malicious activities. Trend Micro's interactive map identifies the prevalence of C&C servers in a given location (Trend Micro 2018).

Arbor Networks is a software company dedicated to network monitoring and network security. The company's products are used to combat DDoS attacks. Currently Arbor Networks' defense software is used by over 90% of all tier 1 Internet service providers globally (Arbor Networks 2018). In recent years, Arbor Networks collaborated with Google Ideas to create an interactive data visualization map that shows the distribution and historical trends of DDoS attacks.

Project Honey Pot is an "antispam company" that identifies and archives malicious IP addresses for security and research purposes (Project Honey Pot 2018). Webmasters simply install the Project Honey Pot software on their website, and the software can parse out the spammers and spambots. Using data gathered from Project Honey Pot, researchers can examine malicious activity associated with specific IP addresses and determine which countries send and receive the highest frequency of spam.

Zone-H falls somewhere between a for-profit enterprise and open-source dataset. It was created in 2002 by a private interest to archive defaced websites. Website defacement is a common form of hacking (Zone-H 2018), defined as the replacement of a website's original content with one's own content. Hackers use Zone-H to brag about their successful defacements. After Zone-H is notified of an attack, the archivists verify it and permanently document the incident in Zone-H's archive. The information on Zone-H is publicly available through its website, but its current owners sell archival data for a negotiated fee.

In 2017 alone, over 1 million website defacements were reported to Zone-H, but most studies only examine the content of the defaced websites. Although criminologists are becoming increasingly more interested in website defacement generally, and the Zone-H archive specifically, the data are still underutilized. Zone-H collects myriad data including the offenders' name and motivation, attack location, system type, and attack domain. Additionally, Zone-H characterizes some website defacements as "special." These special defacements are attacks on "important websites," which are almost always government affiliated.

Like most datasets, Zone-H has limitations. The data are self-reported, which creates potential self-selection bias: those who choose to report their successful defacements may differ from those who do not choose to report. For example, hackers who post defacements on Zone-H may be building their reputations among peers and therefore less experienced than elite hackers with established reputations. Conversely, newbie hackers who can only deface one or two sites may not bother bragging on Zone-H until their skill level increases. In other words, the defacements stored in Zone-H's archive may not be representative of all website defacers.

The Zone-H data do offer valuable insights into website defacement. The data allow for an examination at both the macro-level and event level. When paired with other datasets, the data can be used to correlate macro-level trends in defacement frequency over time and across countries (i.e., Howell et al. 2019).

Researchers can also examine hackers' motivations. Zone-H requires hackers to self-select their motivation when reporting a defacement from a set of possible responses: political, religious, for fun, or other reasons. Although the motivation variable is interesting, the validity and reliability of this measure are suspect. Specifically, the measure is not exhaustive, nor is it mutually exclusive.

The *Cambridge Cybercrime Centre*, a multidisciplinary initiative, has built a robust cybercrime dataset on three aspects of cybercrime: distributed denial of service attacks (DDoS), web forum discussions, and spam. First, the Centre currently operates 100 censors around the world that record incidents of DDoS attacks.

A DDoS attack is an explicit attempt by a hacker to prevent legitimate users from gaining access to online resources, such as retailers, banks, social media, and entertainment streaming services (Lau et al. 2000). DDoS attacks are becoming more common because even those lacking the skills needed to carry out the attack can simply hire a hacker.

The Cambridge Cybercrime Centre data allow researchers to assess DDoS attacks and victimization patterns. Specifically, researchers can view the attackers' and victims' IP addresses, which indicate an attack's location of origin and target's location. The dataset currently includes over four trillion packets. Stated simplistically, the term packet refers to the data being transmitted from one computer to another.

Second, the Cambridge Cybercrime Centre also scrapes data from Internet forums. They have over 40 million posts that can be analyzed for trends, techniques, and even individual criminal trajectories. By analyzing forum data, researchers may describe what activities hackers engage in and how an individual hacker progresses over time. Macdonald et al. (2015), for example, used the language of hackers to identify potential threats against critical infrastructure. Moreover, Benjamin et al. (2015) developed a way to automate the process of identifying existing threats and vulnerabilities using machine learning technologies and information retrieval techniques. Taken together, the use of available forum data can be used for research and the creation of prevention strategies. However, not all hackers discuss their techniques on forums.

Third, the Cambridge Cybercrime Centre gathers phishing and spam data from a variety of sources. Phishing is an increasingly common phenomenon defined as "a scalable act of deception whereby impersonation is used to obtain information from a target" (Lastdrager 2014). Phishing occurs in three stages: (1) the victim receives an email from the offender, (2) the victim takes the action suggested in the message, and (3) the offender monetizes the stolen information (Hong 2012). Phishing costs Internet users billions of dollars a year (McGrath and Gupta 2008). Although much legal investigation has been conducted on spam and phishing, only a few studies have examined it using criminological theory (Kigerl 2012).

Although Cambridge Cybercrime Centre collects one of the most robust cybercrime datasets available to date, they do not utilize random selection. The nonrandom nature of the data calls the generalizability of the data into question. Stated differently, since not all cybercrimes have an equal chance of inclusion, inferences drawn from the data may not reflect the population.

Open-Source Data

The *Malware Domain List* is an example of an open-source repository, like Zone-H, that accepts reports on active malware around the world (MDL 2013). Anyone can report instances of detected malware to the site, such as cybersecurity professionals, white or gray hat hackers, or even black hat hackers. Like other forms of self-reported data, the data collection is biased toward instances of malware that is

detected. Nevertheless, these data have been used to examine the global distribution of malware (Holt et al. 2018).

Another open-source of data are web forums. Multiple studies have used web forums to better understand hacker behavior. For example, Holt et al. (2012) used forum data to explore the social network of a group of Russian hackers. By analyzing this unique source of open-source intelligence, Holt et al. (2012) determined that only a few skilled hackers are active in the forums in comparison to novice hackers. Additionally, Zhang et al. (2015) examined messages posted in hacker forums to create hacker typologies. These studies, among others, demonstrate that researchers can gather open-source intelligence from forums to study cybercrime.

In addition to forum data, researchers have utilized open-source intelligence from various social media platforms to garner insight into cybercrime perpetration. More specifically, Maimon et al. (2017) gathered social media data for a large number of hackers found on the Zone-H archive. Results from their study indicate that social media presence increases attack frequency. Additionally, Babko-Malaya et al. (2017) found that social media behavior varies based on hackers' motivations. Specifically, skilled hackers tend to discuss more technical topics, whereas those motivated by ideology use social media as a recruitment tool.

Conclusion

In this chapter, we have documented various sources of cybercrime data to help guide future research endeavors. The list of analyzable datasets we provided is in no way exhaustive. Similarly, we did not provide a robust discussion of all forms of cybercrime. Therefore, we focused most of our attention on datasets associated with hacking and to a lesser degree online fraud. Rather than a catalog of sources, we also attempted to describe what research has accomplished with these data on specific crimes and to discuss the strengths and limitations of their use.

The data discussed throughout the chapter are gathered from a variety of sources including the FBI, Cambridge Cybercrime Centre, Zone-H, various cybersecurity companies, and several other websites and platforms. All these data allow researchers the opportunity to assess cybercrime correlates of engagement, victimization patterns, and macro-level trends. However, all these data share one major flaw; they do not allow for the assessment of causation.

Despite the methodological limitations of proprietary and self-reported data, studies conducted using these data serve as foundational work. Nevertheless, future research should begin to collect more reliable data from the field to assess specific forms of cybercrime. Specifically, researchers should prioritize the collection of longitudinal data to allow for causal modeling between validated theoretical constructs and cybercrime perpetration and victimization. Additionally, when possible, researchers should employ randomized experimental designs to eliminate the effects of confounding variables.

One potential source of data to be used in experiments comes from honeypot computers, which act as attractive targets to hackers to lure them in. When successfully

employed, network administrators can use honeypots to divert hackers from sensitive information or log the hackers' activities to test the security and resilience of their system. But researchers can also use information from attacked honeypots to determine the skill level and methods of hackers. In addition, researchers can vary aspects of the honeypot systems to determine which factors might act as a deterrent for hackers. For example, using honeypot computers, Maimon et al. (2014) found that warning banners deter system trespasser behavior. However, the warning banner is less effective when the system trespasser has full administrative privileges (Testa et al. 2017).

Although Maimon and his colleagues (Howell et al. 2017; Maimon et al. 2014; Wilson et al. 2015; Testa et al. 2017) have demonstrated that honeypot computers can be used to garner insight into system trespasser behavior, the data is limited in various ways. Specifically, as pointed out by Bossler (2017), honeypot data is collected at the event level, which may lead researchers to make inaccurate assumptions about the attacker. Additionally, the majority of cyberattacks are automated; therefore, honeypot data may include attacks generated by machines rather than individuals (Bossler 2017).

Experimental and quasi-experimental designs can help criminologists gain a fuller understanding of the causal mechanisms underlying cybercrime and cyber victimization (Maimon and Louderback 2018). Moreover, cybercrime scholars can test how traditional criminological theories operate in cyberspace. Unfortunately, this cannot be easily accomplished through secondary data analysis and convenience samples of college students. Criminologists must trade convenience for innovation and gather data on active offenders. Effective policy stems from good research, which can only be conducted with the use of valid and reliable data.

The data generated by networked computing devices continues to generate massive amounts of data on a multitude of user activities that can be used to track and predict human behavior. This unprecedented tracking of our daily activities brings with it new privacy concerns until recently only imagined in George Orwell's *Nineteen Eighty-Four* Certainly, researchers will need to be mindful of how the data are used to protect human subjects from unanticipated harm. The promise of measuring the interaction between humans and networked systems, however, should help us understand criminal victimization and offending that is not possible for non-cybercrimes. Knowing what one does online, how often, and under what circumstances would help develop existing criminological theories but also may lead to new theories. Ultimately our understanding of crime is only as good as the data we have to test or create theories.

To improve the current state of cybercrime data, researchers should take two important steps: (1) facilitate data consortiums where proprietary, official, and public data are made available for researchers; and (2) explore what data are and can be generated from network users; then make this available. For the first step, businesses and government must work together to create data consortiums where data are managed and uploaded for research use. Many proprietary datasets are kept classified or confidential because cybersecurity firms or government agencies feel the need to protect the names of clients or keep secret cybersecurity failures. This is

certainly understandable. However, data can always be made safe for public use by de-identifying the cases or aggregating the information. Criminologists routinely use information about individual victims, prisoners, or agency activities without divulging any sensitive details. This can certainly be done for cybercrime data. It took an act of Congress in 1930 to create the UCR; hopefully, affected businesses and agencies can see the need for collaboration for their own benefit.

The second step requires interdisciplinary research between social scientists and computer and software engineers. As computing devices become more interconnected through software applications, social media, and Internet of things, we should be able to use this vast array of data to understand how humans interact with each other and through technology. We know that human perceptions can be manipulated through software algorithms that use online activity to predict one's shopping preferences to political ideology. By understanding what data are being created, we might be able to find correlates of offending and victimization that go beyond survey questions. Social media use (amount not content), for example, might be predictive of psychological issues (depression) or in the aggregate geographic disparities (anomie).

In conclusion, data that criminologists can use to understand cybercrime is like that used for traditional crime. The main difference is that cybercrimes, like traditional fraud, remain underreported even within self-reported surveys. Nevertheless, we have discussed several promising sources in this chapter. We believe the data generated by computers and networks (including cellular) offers promise for helping criminologists explain cybercrime through existing theory or to develop new theory to explain the growing problem of cybercrime.

References

Arbor Networks. (2018). *Insight into the global threat* landscape. Retrieved from https://www.netscout.com/report

Babko-Malaya, O., Cathey, R., Hinton, S., Maimon, D., & Gladkova, T. (2017). Detection of hacking behaviors and communication patterns on social media. In *2017 IEEE international conference on big data (Big Data)* (pp. 4636–4641). Boston, MA: IEEE.

Benjamin, V., Li, W., Holt, T., & Chen, H. (2015). Exploring threats and vulnerabilities in hacker web: Forums, IRC and carding shops. In *Intelligence and Security Informatics (ISI), 2015 IEEE international conference on* (pp. 85–90). Baltimore, MD: IEEE.

Bitdefender. (2018). *Cyberthreat real time map*. Retrieved from https://threatmap.bitdefender.com

Bossler, A. M. (2017). Need for debate on the implications of honeypot data for restrictive deterrence policies in cyberspace. *Criminology & Public Policy, 16*(3), 681–688.

Button, M., Nicholls, C. M., Kerr, J., & Owen, R. (2014). Online frauds: Learning from victims why they fall for these scams. *Australian & New Zealand Journal of Criminology, 47*(3), 391–408.

Federal Bureau of Investigation. (2019a). *Crime in the United States 2018*. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/about-cius

Federal Bureau of Investigation. (2019b). *National incident-based reporting system*. Retrieved from https://www.fbi.gov/services/cjis/ucr/nibrs

Federal Bureau of Investigation Internet Crime Complain Center (IC3). (2017). *2017 Internet crime report.* Retrieved from https://pdf.ic3.gov/2017_IC3Report.pdf

Holt, T. J., Strumsky, D., Smirnova, O., & Kilger, M. (2012). Examining the social networks of malware writers and hackers. *International Journal of Cyber Criminology, 6*(1).

Holt, T. J., Burruss, G. W., & Bossler, A. M. (2018). Assessing the macro-level correlates of malware infections using a routine activities framework. *International Journal of Offender Therapy and Comparative Criminology, 62*(6), 1720–1741.

Hong, J. (2012). The state of phishing attacks. *Communications of the ACM, 55*(1), 74–81.

Howell, C. J., Maimon, D., Cochran, J. K., Jones, H. M., & Powers, R. A. (2017). System trespasser behavior after exposure to warning messages at a Chinese computer network: An Examination. *International Journal of Cyber Criminology, 11*(1), 63–77.

Howell, C. J., Burruss, B. W., Maimon, D., & Sahani, S. (2019). Website defacement and routine activities: Considering the importance of hackers' valuations of potential targets. *Journal of Crime and Justice, 42*, 536.

Huff, R., Desilets, C., & Kane, J. (2010). *National public survey on white collar crime*. Fairmont: National White Collar Crime Center.

Kaspersky Lab. (2018). *Cyberthreat real-time map.* Retrieved from https://cybermap.kaspersky.com

Kigerl, A. (2012). Routine activity theory and the determinants of high cybercrime countries. *Social Science Computer Review, 30*(4), 470–486.

Lastdrager, E. E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science, 3*(1), 1–10.

Lau, F., Rubin, S. H., Smith, M. H., & Trajkovic, L. (2000). Distributed denial of service attacks. In *Systems, man, and cybernetics, 2000 IEEE international conference on* (Vol. 3, pp. 2275–2280). Nashville, TN: IEEE.

Lee, M., Pak, S., Lee, D., & Schapiro, A. (1999). Electronic commerce, hackers, and the search for legitimacy: A regulatory proposal. *Berkeley Technology Law Journal, 14*(2), 839.

Macdonald, M., Frank, R., Mei, J., & Monk, B. (2015). Identifying digital threats in a hacker web forum. In *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015* (pp. 926–933). Paris, France: ACM.

Maimon, D., & Louderback, E. R. (2018). Cyber-dependent crimes: An interdisciplinary review. *Annual Review of Criminology*, (0), 191–216.

Maimon, D., Alper, M., Sobesto, B., & Cukier, M. (2014). Restrictive deterrent effects of a warning banner in an attacked computer system. *Criminology, 52*(1), 33–59.

Maimon, D., Fukuda, A., Hinton, S., Babko-Malaya, O., & Cathey, R. (2017). On the relevance of social media platforms in predicting the volume and patterns of web defacement attacks. In *2017 IEEE international conference on big data (Big Data)* (pp. 4668–4673). Boston, MA: IEEE.

Malware Domain List (MDL). (2013). *Malware domain list frequent asked questions*. Retrieved from http://www.malwaredomainlist.com

McAfee. (2018). *Global virus map.* Retrieved from https://home.mcafee.com/virusinfo/global-virus-map

McGrath, D. K., & Gupta, M. (2008). Behind phishing: An examination of phisher modi operandi. *LEET, 8*, 4.

Muniz, C. N. (2019). *Sexual assault and robbery disclosure: An examination of Black's theory of the behavior of law* (Doctoral dissertation, University of South Florida).

Pangaria, M., & Shrivastava, V. (2013). Need of ethical hacking in online world. *International Journal of Science and Research (IJSR),* India Online. ISSN: 2319–7064, 529–531.

Project Honey pot. (2018). *Project honey pot*. Retrieved from https://www.projecthoneypot.org

Rantala, R. (2008). Cybercrime against businesses, 2005. Retrieved from http://www.bjs.gov/index.cfm?ty=pbdetail&iid=769

Reyns, B. W. (2015). A routine activity perspective on online victimisation: Results from the Canadian general social survey. *Journal of Financial Crime, 22*(4), 396–411.

Reyns, B. W., Henson, B., Fisher, B. S., Fox, K. A., & Nobles, M. R. (2016). A gendered lifestyle-routine activity approach to explaining stalking victimization in Canada. *Journal of Interpersonal Violence, 31*(9), 1719–1743.

Schafer, J. A., Lee, C., Burruss, G. W., & Giblin, M. J. (2018). College student perceptions of campus safety initiatives. *Criminal Justice Policy Review, 29*(4), 319–340.

Testa, A., Maimon, D., Sobesto, B., & Cukier, M. (2017). Illegal roaming and file manipulation on target computers: Assessing the effect of sanction threats on system trespassers' online behaviors. *Criminology & Public Policy, 16*(3), 689–726.

Trend Micro. (2018). *Global botnet map*. Retrieved from https://botnet-cd.trendmicro.com

U.K. Office for National Statistics. (2019) Crime and justice. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice

Wilson, T., Maimon, D., Sobesto, B., & Cukier, M. (2015). The effect of a surveillance banner in an attacked computer system: Additional evidence for the relevance of restrictive deterrence in cyberspace. *Journal of Research in Crime and Delinquency, 52*(6), 829–855.

Zhang, X., Tsang, A., Yue, W. T., & Chau, M. (2015). The classification of hackers by knowledge exchange behaviors. *Information Systems Frontiers, 17*(6), 1239–1251.

Zone-H. (2018). *Unrestricted information.* Retrieved from http://www.zone-h.org