Online Crime Monitoring

David Décary-Hétu, Centre international de criminologie comparée (CICC), Université de Montréal

Abstract

The aim of this chapter is to explore the research opportunities afforded by the Internet. To do so, this chapter will describe both the indirect and direct monitoring techniques used to monitor the traces offenders leave online. Three case studies, on prostitution, financial fraud and online drug dealing, will demonstrate the potential of online monitoring. The online monitoring of offenders, both indirect and direct, arguably offers one of the most interesting opportunities for criminologists and forensic scientists to collaborate with one another. Forensic scientists are increasingly developing their expertise in the collection and analysis of digital traces. Criminologists, on the other end, have developed methods and research design to understand offending behaviour even when very limited data are available. Through online monitoring, criminologists are likely to gain a much deeper insight into offending trajectories, including who offenders are, how they become offenders, how they operate and how they desist from crime. Using data collected by forensic scientists, criminologists may be able to answer—or shine a new light on—questions that have plagued criminology for quite some time.

Traces and Offenders on the Internet

Online communities have been using computers and networking equipment since the 1970s. At first, the communities were mainly geared toward academic and military applications and allowed for the rapid dissemination of information (Salus and Vinton 1995). Access to online networks was restricted to a select few, given the scarcity of computer resources. As technology democratized in the 1980s, new types of communities emerged. These communities were much more open, in many cases granting access to anyone and everyone. They took on the forms of virtual bulletin boards (Aboba 1993; Parks and Floyd 1996) and allowed participants to phone in to remote computers to leave

messages for one another and to exchange files. The networks that hosted these first-generation online communities were very slow compared to today's standards. Their speed was limited by the phone lines and modems they used, which were at first limited to 56 kpbs, about 0.1 percent of the speed of an average modern Internet connection (Speedtest 2016). These networks also had very limited concurrent connections since each user needed to have an open phone line to connect to. Finally, the networks were expensive since some had subscriptions costs and were in different locales from their users who consequently had to pay long-distance charges for the duration of their call. These limits helped to push online community participants in the 1990s to a new type of network, the Internet. The Internet promised higher speeds, a greater number of concurrent connections on a single system and no long-distance charge when connecting to systems anywhere in the world.

Internet communities quickly grew in size and have taken many forms that can generally be classified into one of four groups. Discussion forums (Guzdial and Turns 2000) are asynchronous communication platforms akin to the old virtual bulletin boards. Users leave messages for each other and can exchange files. The forums keep a history of their participants' activities, which can be analyzed at a later date. Chat rooms (Décary-Hétu et al. 2014; Reid 1991) are similar to discussion forums in that they allow their participants to exchange messages and files but offer only synchronous communications. These communications are usually not stored, which means that participants who are away when a message is broadcast are unlikely to be able to read it later. Blogs (Boulos, Maramba and Wheeler 2006) are websites where authorized users can post messages. Other participants can comment on this information but are usually not able to post new information. Finally, online markets (Turban et al. 2009) are virtual marketplaces where one or more vendors can put up products and services for sale. Potential customers can compare ads and then select the product that best fits their need and the vendor with the most trustworthy history. Amazon and eBay are two examples of successful online markets that have endured the test of time.

Online communities make it easy for individuals to network with one another and to buy and sell products and services. The technologies they use are in many cases free and easy to learn, meaning that everyone can adapt them to their needs. It is therefore not surprising that past research has provided many examples of how all of the online communities' platforms, going back to the 1980s, were at some point adopted by offenders. Virtual bulletin board systems were very popular with phone phreaks who sought to manipulate the telephone networks and play tricks on one another (Meyer 1989). They were also an essential part of the intellectual property fraud community that sought to illegally share proprietary content such as movies and software for free (Tickle 1994). Later on, chat rooms were used extensively to share once again intellectual property such as music and movies (Cooper and Harrison 2001; Honick 2005). Many cases were also reported where chat rooms were used to buy and sell personal and financial information as well as malware (Benjamin et al. 2015; Franklin et al. 2007; McCarty 2003). Discussion forums were also used as online illicit markets where offenders could post ads for the illicit goods and services for sale and receive inquiries from potential customers (Décary-Hétu and Laferrière 2015). More recently, second-generation online illicit markets named cryptomarkets have appeared and offer the ability to buy and sell illicit goods and services with a high level of impunity (Martin 2014). These cryptomarkets take advantage of the Tor network, a subset of the Internet where the IP address of users and websites (their unique identifier online) is obfuscated using a series of anonymous relays. Tracking their connection and location is therefore very difficult. Cryptomarkets also make use of virtual currencies like bitcoin, which allow for instantaneous, international and anonymous transactions where all the parties involved are known only by a unique identifier that is hard to associate with a real identity.

The Internet and its online communities have provided offenders with many new opportunities and benefits. First, they have created new and international convergence settings where offenders can meet and share information (Soudijn and Zegers 2012). These have exposed offenders to the best practices of their peers and helped them to maximize their benefits while limiting their risks. The Internet, through its multiple online communities, also allowed for the creation of online markets which facilitate the

sale of illicit goods and services (Décary-Hétu 2013). Cryptomarkets alone are believed to be facilitating the sale of hundreds of millions in US dollars of illicit goods and services per year and have shown just how big the underground economy has become (Soska and Christin 2015). The adaptation of offenders to the Internet may at first be seen as a challenge for regulators and law enforcement agencies. This chapter argues, however, that it also represents an opportunity for researchers who study offenders.

Indeed, no matter which community offenders are involved in, it is very likely that their activities will leave traces that can be collected at a later time. Most communities keep logs of their participants' activities for quality of service purposes or to ensure that they can deliver the service they promised. Discussion forums, for example, need to keep a copy of all of their members' messages if they are to provide them with a history of the forum activities. Additional information about members is also routinely collected. Discussion forums will often log the IP addresses of their members so that they can more effectively block the members who have been banned from the forum. Third parties also collect much of the information available in online communities. The Internet Archive (Internet Archive 2017) crawls the entire Internet, seeking to preserve a historical copy of its content. Doing so, it visits online communities that bring together offenders, and traces of their activities can therefore be collected there. The cache of web pages stored by the Google search engine provides an alternative means of accessing traces left by offenders online.

The aim of this chapter is to explore the research opportunities afforded by the Internet. To do so, this chapter will describe both the indirect and direct monitoring techniques used to monitor the traces offenders leave online. It will then provide examples of how these traces were used to better understand three type of offenders: those involved in prostitution, in financial fraud or in drug dealing. This chapter will conclude with a discussion of the challenges of collecting and analyzing traces online and the future developments we should expect from this field of research.

Indirect and Direct Online Monitoring of Traces on the Internet

To take advantage of the traces left online by offenders, indirect and direct monitoring techniques can be used. Indirect monitoring refers to the use of leaked documents posted on the Internet. While there are many reasons why documents may leak online (e.g., a non-secure server allowing Google to index confidential information), the most interesting leaks have so far been a result of the fierce competition between online offenders. Offenders may compete with each other online to prove their worth to others, for revenge and, perhaps more importantly, to take down competitors. Offenders who run online markets earn a commission on the sales that their platform facilitates (Christin 2013). They may also collect membership fees (Yip et al. 2013), a per-listing fee (Hutchings and Holt 2014) or fees for other services such as escrow services (Holt 2013). As such, offenders may have a vested interest in driving the sale of illicit goods and services toward their platform to increase their commissions. One way to achieve this is to destroy competitors' reputations by hacking them and distributing the confidential information of their communities' participants online (Poulsen 2012). Such actions have led to the publication of backups of databases of online forums and customers lists that contain a wide array of information on offenders (Motoyama et al. 2011). These leaks can contain gigabytes of information and include public and private messages, email addresses, IP addresses and passwords. In other cases, offenders use a more targeted attack, also called doxing, which is the act of divulging the identity of an individual online (Leong and Morando 2015). The information leaked in those cases is much smaller but may provide very detailed reports on the characteristics and identity of offenders.

Indirect monitoring is time sensitive. Leaks are often short-lived and posted on websites that may not remain online for long. It is therefore difficult to find files that were leaked years or even months in the past. Constantly monitoring the Internet for leaks is therefore essential to collecting traces on offenders. Researchers should build online sentinel software that looks for keywords in online forums, specialized blogs and social media networks. Further information could also be gleaned from security podcasts, which often review important leaks. Indirect monitoring can be used to extract information from third parties that collect data on the Internet in general. Search engines and the Internet

Archive, for example, constantly crawl the open Internet, looking for content to index and store. Such repositories can be used to extract information from offenders without ever getting in direct contact with them. This approach has the added benefit of stealth and limiting the exposure of researchers. There have been cases when online offenders have sought to strike back at researchers who published too much information about them (Krebs 2016), and adding this layer of abstraction between the researcher and the offenders may help to protect the former from retaliation (see Barratt and Maddox [2016] for further discussion on this topic).

Rather than wait and hope to find interesting leaks, researchers can adopt a more proactive stance and monitor the online activities of offenders themselves (Décary-Hétu and Aldridge 2015). The technology used to conduct such monitoring will vary depending on the platforms that need to be monitored. Chat rooms have been monitored in the past using automated systems but have proven to be challenging (Décary-Hétu et al. 2014) since the most interesting discussions and activities happen in password-protected chat rooms. Many chat rooms also tend to ban users who do not actively participate in the activities of the community either by posting messages or buying and selling goods and services. This active participation is often prohibited by ethics review, and the obstacles to maintaining a stealth presence on chat rooms may explain the scant research that has been published using chat rooms as data sources. Much more research has used the mirroring technique, which uses a software package to collect web pages (Aldridge and Décary-Hétu 2016; Chen 2011; Soska and Christin 2015; Westlake et al. 2012). The software is a two-stage process that starts with a crawler robot that connects to a server and downloads a web page. Once downloaded, the robot stores the web page for later and extracts from it all of the hyperlinks. It then requests each hyperlink, looking for more links to put in its download queue. This process is repeated iteratively until the download queue is empty. At that point, the software switches to its scraper robot, which goes through each of the downloaded pages, looking for specific information.[1] Researchers have to teach the scraper robot what to look for and to store the information it has extracted in structured databases. The speed at which information can be collected

will vary depending on the available resources, but a simple crawler robot can easily download hundreds of pages per minute.

A small portion of the Internet is indexed by search engines and easily reachable, simply by tapping a URL in a browser (Bindal and Muktawat 2010). Most crawlers share the same limit and are unable to reach the deep and darknet. The deep net is the part of the Internet that is protected by access control systems. It includes, for example, the salary information of companies, information deemed confidential. Some of the deep net can be accessed by registering accounts on websites and then passing the credentials to the crawler so that it can index the content protected by the login page. However, this requires a more sophisticated crawler that may be more difficult to design. Many login forms are also protected by CAPTCHAs and questions that are difficult to answer via automated crawlers. These can, however, be bypassed through the use of grey market commercial services that employ thousands of humans paid a few pennies to solve the CAPTCHAs and the questions (Petsas et al. 2015). The darknet is a subset of the deep web and refers to the section of the Internet that requires that all communications be encrypted. The darknet is most commonly known for hosting the Tor network but also encompasses other networks such as the I2P network and the Freenet network (Conrad and Shirazi 2014). Crawlers can be configured to index the content of websites hosted on the darknet relatively easily, but discovering websites to crawl may be more difficult. Mainstream search engines have mostly stayed away from the darknet websites for now and some queries that take advantage of the TOR2WEB bridge to the Tor network may return some results by using search terms like ‘site:onion.link illicit content’. Some Tor specific search engines can also be found but their reliability is still mostly unknown. Researchers will therefore need to build a list of websites without much help from traditional search tools and look at discussion forums, blogs and social network posts.

Social networks like Facebook, Twitter, Instagram and Snapchat are also themselves very rich sources of data for researchers and need be indexed actively as leaks are few and far between. With its 1.6 billion users (Adler 2016), Facebook is without contest the largest social network service, and Americans spend on average five hours a week on social

networks in general (Mosendz 2017). Given the proportion of the world population that uses social networks, it is to be expected that offenders will also use social networks, to target victims and to post messages about themselves. Social networks have become very adept at detecting crawlers and blocking them from indexing their content. They do, however, offer access to their data servers through their application program interfaces (APIs). APIs are formalized rules that guide how researchers can access information from social networks (Graham 2008). They commonly limit what information can be collected and the number of requests that can be made on a daily or hourly basis. Some social networks like Twitter are known for their very open APIs while others like Instagram impose very strict limits on what can be collected through their APIs as well as who is able to access them. Many social networks are open to commercial contracts where researchers can negotiate for an enhanced access to the social networks data in exchange for money or other compensation.

The direct monitoring of offenders online requires a level of computer technical skills that most researchers outside the computer science world are unlikely to have. Many commercial services have been launched to help these researchers—although they mostly focus on corporate customers—to actively monitor online activities (see, for example, Import.io [2017]; Scraping Hub [2017]); and Web Scraper [2017]). These commercial services are costly and will still require that the researchers invest serious resources into the cleaning and formatting of the data they collect.

Exploiting Traces Collected on the Internet

Given the costs of hiring commercial services to actively monitor offenders online, researchers may also resort to collecting their data by hand, building their databases one web page at a time. This is the technique that Holt and his team (Blevins and Holt 2009; Holt and Blevins 2007; Holt, Blevins and Kuhns 2008, 2014) opted to use when monitoring the online activities of sex workers and their johns. Most work focusing on prostitution had up to then used police data and interviews to understand the motivation, practices and business of prostitution. There are well-known limits to using police data since they are severely affected by the dark figure of crimes, an issue even more

prevalent with consensual offences (MacDonald 2002). Interviews are for their part likely to be limited in size and scope and are therefore unlikely to provide a generalized understanding of prostitution. Turning to the online monitoring of offenders allows for access to untainted data since johns and sex workers communicating online are unlikely to change how they talk and behave based on the remote possibility that their messages may at some point be used as research material. Of course, not all actors involved in prostitution will use the Internet to communicate, but given the pervasiveness of computer-mediated communications in everyday lives, it is likely that a fair share of actors will at some point be caught in online monitoring of their activities. Using this monitoring, Holt found that discussion forums actually played an important role in finding sex workers through their online ads and references from other johns. These johns used argot to communicate with one another, which created a somewhat hermetic community around them, insulating them from unknown actors and, to a certain extent, law enforcement officers. Risk management is a big concern for johns given the reputational—and criminal—risks associated with hiring sex workers. Many online discussions would deal with the regulating operations of law enforcement agencies and the identification of fake ads trying to lure johns to law enforcement officers posing as sex workers. Online monitoring gave Holt and his team a view into the adaptation of johns and the techniques they used to displace their activities and reduce their risks. These techniques also applied to sex workers themselves, who sometimes opted to steal from and assault their clients.

Dubrawski et al. (2015) have suggested that it is possible to monitor sex workers online and that this monitoring can generate interesting results. Dubrawski et al. collected massive numbers of sex workers' online ads and identified predictors of listings that could feature trafficked women. Contact information of sex workers can also be used to build networks of providers of sex services online, leading to better intelligence about the state of the industry. Alvari et al. (2016) and Nagpal et al. (2016) also provide support for using machine learning to identify sex workers' ads potentially linked to human trafficking, although all research stops short of actually contacting the sex workers to confirm whether or not they are victims of human trafficking. This represents an

interesting vector for forensic science to identify ads of interest and for criminologists to build on this research to find research subjects.

An important perk of using the online monitoring of offenders is that researchers can glean a direct understanding of offenders from the point of view of offenders themselves. This was the case for sex workers and it is also the case for financial fraud. Financial fraudsters are still using physical means such as mail theft to steal credit cards and make purchases in stores that are later resold on the Internet (Peretti 2008). Fraudsters have also, however, moved to the Internet to steal financial information (Décary-Hétu and Leppänen 2013). This can be achieved through a number of means including hacking payment processing systems and luring unsuspecting users to fake websites that steal their bank credentials (Gomzin 2014; James 2005). Offenders involved in the theft of financial information are in large part unwilling or unable to monetize this information and instead opt to sell the information to other fraudsters who have mastered the art of monetizing stolen financial information. This can be achieved by printing fake credit cards and making in-store purchases which are later refunded for cash or more commonly resold online (Meijerink 2013). Another technique used is to make online purchases and once again have the goods resold online (Hutchings and Holt 2014). To bring these two types of financial fraudsters together, online markets were launched, first on discussion forums and in chat rooms and later on cryptomarkets. Past research has shown that very large communities with thousands of financial fraudsters could be found online (Motoyama et al. 2011). They are hosted on online markets that facilitate the sale of financial information and its subsequent monetization, first by matching vendors and customers, and second by sharing the best practices developed by offenders. Fraudsters looking to buy financial information can look at the past history of vendors, including the feedback from past customers, to determine the trustworthiness of vendors (Yip et al. 2013). Prices can also be easily compared to find the right product at the right price. There appear to be very large quantities of financial information for sale on online markets and its price appears to be low, with the price of a single credit card number at less than $10. Some researchers believe that these large quantities of financial information are finding buyers and that the impact of financial fraudsters on the financial

system is significant (Holt, Smirnova and Chua 2016). Others have argued that the low price for the financial information is indicative of its low value and that much of the information available on online markets is either fake—and therefore of no value—or never sold (Herley and Florêncio 2010). The impact of financial fraud would therefore be much lower.

Large-scale datasets have also been collected on a new and innovative use of online markets to sell illicit drugs (Aldridge and Décary-Hétu 2016; Soska and Christin 2015). These sales are mostly concentrated on cryptomarkets where dealers put up listings for specific amounts of drugs. Customers can once again study the history of dealers and compare prices (Martin 2014). The drugs are shipped stealthily through the mail to the customer, who is unlikely to be arrested as long as she purchases from dealers who are located in the same country as she is. Parcel inspections are indeed unlikely when the parcel does not cross national borders. Many researchers have sought to understand the displacement of drug dealing from the physical to the virtual world by monitoring the online activities of drug dealers and drug users. Their research has indicated that the size and scope of cryptomarkets has vastly increased over the past few years, with annual sales ranging now in the hundreds of millions of dollars (Soska and Christin 2015). Cryptomarkets now facilitate the sale of hundreds of types of drugs with dozens of countries being involved in their sale (Kruithof et al. 2016). Major Western countries like the United States of America, the United Kingdom and the Netherlands are major suppliers of drugs online. While most listings are intended either for drug users or social suppliers, many listings are also geared toward other drug dealers who may be looking to source their stock online. Aldridge and Décary-Hétu (2016) found that such dealers generated a significant portion of revenues and that cryptomarkets were to some extent business-to-business platforms. The direct online monitoring of offenders has therefore allowed researchers to understand the size and scope of online drug dealing. It has also allowed for the understanding of the impact of police operations on drug markets (Décary-Hétu and Giommoni 2016; Van Buskirk et al. 2017). In line with past research on traditional drug markets, cryptomarket research found that the police operations had little impact on the drug dealing activities and that targeting the cryptomarkets

themselves without making dealer arrests simply disrupted the dealers' activities for a limited period of time. While not surprising, the results based on cryptomarket data offers a much more conclusive picture of the impact of police operations by analyzing all of the drug market activities rather than a subset generated through select interviews of drug dealers and drug users. There are, finally, many more examples of physical and digital traces that can be collected on cryptomarkets (see 'Combining Physical and Digital Traces' in this volume). These include using the PGP keys of vendors to link their accounts together across cryptomarkets and the analysis of physical drugs bought on cryptomarkets.

Challenges and Future Developments

Forensic science focuses on the study of traces that are generated through criminal activities. The online monitoring of offenders, both indirect and direct, offers arguably one of the most interesting opportunities for criminologists and forensic scientists to collaborate with one another. Forensic scientists are increasingly developing their expertise in the collection and analysis of digital traces. These traces can either be artefacts of a user on a desktop computer or traces left by offenders on the Internet. With their hard science background, forensic scientists have already developed some of the skills and methodology necessary for the collection of online traces. Criminologists, on the other end, have developed methods and research design to understand offending behaviour even when very limited data were available. This was the case for sex worker research that had to be based on interviews and police data reports. Through online monitoring, criminologists are likely to gain a much deeper insight into offending trajectories including who offenders are, how they become offenders, how they operate and how they desist from crime. Using data collected by forensic scientists, criminologists may be able to answer—or shine a new light on—questions that have plagued criminology for quite some time.

A major challenge for forensic scientists and criminologists will be to find a way to manage the massive amount of traces that offenders generate every day. On cryptomarkets alone, thousands of offenders offer hundreds of thousands of products and

services on a daily basis (Kruithof et al. 2016). These generate thousands of sales every day. Collecting and managing this information will require important investments in computer equipment but also the development of optimizing strategies to store and search all of this information. Another concern is with the validity of the data that is collected. It is easy, for example, for cryptomarket administrators to generate fake reviews to make potential customers believe that their cryptomarket is very active when that is not the case. Researchers who use the feedback as proxies for the level of activity of cryptomarkets may falsely claim that a cryptomarket is gaining traction when, in fact, the activity is only a mirage designed by the cryptomarket administrators. Much effort will also need to be invested in the cleaning, structuring and validation of the data as debates have already started about the validity of certain data collection efforts (Munksgaard et al. 2016).

One strategy to validate information may be to use a small data approach rather than a big data approach. The online monitoring of offenders can generate very rich datasets where offenders post long and detailed messages about their experience. By concentrating on more qualitative analyses, researchers are likely to be better able to detect suspect postings and to discard them from their research. They are also more likely to derive generalizable results through effective sampling of the massive amount of information that is collected. Big data quantitative analyses will of course be needed, if only to measure the size and scope of the online offending communities. But the focus of research should perhaps stay on small and more detailed analyses rather than bulk analyses.

Another challenge for researchers will be to gain access to the deep and darknet. Much of the Internet's information is located in the deep and darknet and these have remained largely untouched by researchers so far. Netnographies (Kozinets 2010)—ethnographies that use the Internet—can help researchers gain acceptance within offenders' communities and thereafter access more private sections. But these actions will take time and only deliver results in the long term. Ethics review boards strictly control how information can be collected online and the hacking of closed communities is still rightly

out of reach of researchers. As such, researchers should be cautious when seeking to publish generalizable results. It is also very likely that more professional offenders are communicating with each other over private channels that are difficult for researchers to monitor. The offenders that are commonly included in studies may therefore be considered amateurs rather than established professional offenders.

Social networks sites like Facebook and Snapchat will also represent a major challenge for researchers. While many of them offer access to their data through APIs, some do not, and others limit the speed at which information can be gathered. It would be interesting, for example, to evaluate the risk of identity theft of citizens in a given country by downloading, for all of the citizens, the available personal identifiable information they have posted on social networks. Such an experiment would likely require that millions of requests be made to a social networks' API and, with the current rate limiting of APIs, this data collection could take decades to complete. As individuals move more and more of their life online (and to social networks), gaining access to social networks will be a challenge that researchers need to address. At this point, examples of collaboration between researchers and social network sites are rare.

Moving forward, researchers will need to address the representativeness of online offenders. This chapter argues that crime is transitioning more and more to the Internet but such crime will likely never replace crime in the physical world. Indeed, some forms of violence, for example, can only occur in the physical world. Still, online harassers and threat makers do use the Internet and thus some broad knowledge about violence could still be gleaned through the online monitoring of offenders. To what extent this knowledge would apply to offline offenders is still unknown, however. Exploring this question should be included in future research designs.

A non-exhaustive study of the programs of major forensic and criminological association conferences shows that the online monitoring of offenders is still a data collection method that is seldom used in either discipline. This chapter has highlighted the potential of this method using three case studies where the online monitoring of offenders led to

interesting and innovative research. The lack of enthusiasm for the online monitoring of offenders could be explained by the computer skills that are needed to build data collection infrastructures. These often require large investments in economic and human resources to design, launch and run. Commercial services are available to help those that lack the proper computer skills but their existence is still a well-kept secret in the research field. Researchers may also lack the background necessary to formulate problems and to evaluate the quality of the propositions of commercial services. The aim of this chapter has been to provide an overview of how online monitoring of offenders could be achieved and to demonstrate its ability to push back the limits both of forensic science and criminology. As the Internet becomes more and more ingrained in all our lives, the online monitoring of offenders is likely to become an essential part of both forensic science and criminology.

References

Aboba, B. 1993. *The Online User's Encyclopedia: Bulletin Boards and Beyond*. Boston: Addison-Wesley.

Adler, E. 2016. 'Social Media Engagement: The Surprising Facts about How Much Time People Spend on the Major Social Networks'. Business Insider. http://www.businessinsider.com/social-media-engagement-statistics-2013-12.

Aldridge, J., and D. Décary-Hétu. 2016. 'Hidden Wholesale: The Drug Diffusing Capacity of Online Drug Cryptomarkets'. *International Journal of Drug Policy* 35(September): 7–15.

Alvari, H., P. Shakarian and J.K. Snyder. 2016. 'A Non-Parametric Learning Approach to Identify Online Human Trafficking'. In 2016 IEEE Conference on Intelligence and Security Informatics (ISI), Tucson, AZ: 133–138.

Barratt, M.J., and A. Maddox. 2016. 'Active Engagement with Stigmatised Communities through Digital Ethnography'. *Qualitative Research* 16(6): 701–719.

Benjamin, V., W. Li, T. Holt and H. Chen. 2015. 'Exploring Threats and Vulnerabilities in Hacker Web: Forums, IRC and Carding Shops'. Paper presented at the Intelligence and Security Informatics (ISI), 2015 IEEE International Conference, Baltimore, MD.

Bindal, S., and H.S. Muktawat. 2010. *Deep Web*. Online: https://www.researchgate.net/publication/261773660_Deep_Web.

Blevins, K.R., and T.J. Holt. 2009. 'Examining the Virtual Subculture of Johns'. *Journal of Contemporary Ethnography* 38(5): 619–648.

Boulos, M.N.K., I. Maramba and S. Wheeler. 2006. 'Wikis, Blogs and Podcasts: A New Generation of Web-Based Tools for Virtual Collaborative Clinical Practice and Education'. *BMC Medical Education* 6(1): 41.

Chen, H. 2011. *Dark Web: Exploring and Data Mining the Dark Side of the Web*. Vol. 30. New York: Springer Science & Business Media.

Christin, N. 2013. 'Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace'. Paper presented at the Proceedings of the 22nd International Conference on the World Wide Web, Rio de Janeiro.

Conrad, B., and F. Shirazi. 2014. 'A Survey on Tor and I2P'. Paper presented at the Ninth International Conference on Internet Monitoring and Protection (ICIMP2014), Paris.

Cooper, J., and D.M. Harrison. 2001. 'The Social Organization of Audio Piracy on the Internet'. *Media, Culture & Society* 23(1): 71–89.

Décary-Hétu, D. 2013. 'Le capital virtuel: entre compétition, survie et réputation'. PhD Diss., Université de Montréal.

Décary-Hétu, D., and J. Aldridge. 2015. 'Sifting through the Net: Monitoring of Online Offenders by Researchers'. *European Review of Organised Crime* 2(2): 122–141.

Décary-Hétu, D., B. Dupont and F. Fortin. 2014. 'Policing the Hackers by Hacking Them: Studying Online Deviants in IRC Chat Rooms'. In A.J. Masys, ed., *Networks and Network Analysis for Defence and Security*, 63–82. New York: Springer.

Décary-Hétu, D., and L. Giommoni. 2016. 'Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Ononymous'. *Crime, Law and Social Change* 67(1): 55–75.

Décary-Hétu, D., and D. Laferrière. 2015. 'Discrediting Vendors in Online Criminal Markets: Disrupting Criminal Networks' In A. Malm and G. Bichler, eds, *Disrupting Criminal Networks: Network Analysis in Crime Prevention*, 129–152. Boulder: Lynne Reinner.

Décary-Hétu, D., and A. Leppänen. 2013. 'Criminals and Signals: An Assessment of Criminal Performance in the Carding Underworld'. *Security Journal* 29(3): 442–460.

Dubrawski, A., K. Miller and M. Barnes. 2015. 'Leveraging Publicly Available Data to Discern Patterns of Human-Trafficking Activity'. *Journal of Human Trafficking* 1(1): 65–85.

Franklin, J., A. Perrig, V. Paxson and S. Savage. 2007. 'An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants'. Paper presented at the ACM Conference on Computer and Communications Security, Chicago.

Gomzin, S. 2014. *Hacking Point of Sale: Payment Application Secrets, Threats, and Solutions*. Indianapolis: John Wiley & Sons.

Graham, W. 2008. *Facebook API Developers Guide*. New York: Infobase Publishing.

Guzdial, M., and J. Turns. 2000. 'Effective Discussion through a Computer-Mediated Anchored Forum'. *The Journal of the Learning Sciences* 9(4): 437–469.

Herley, C., and D. Florêncio. 2010. 'Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy.' In T. Moore, D. Pym and C. Ioannidis, eds., *Economics of Information Security and Privacy*, 33–53. New York: Springer.

Holt, T.J. 2013. 'Exploring the Social Organisation and Structure of Stolen Data Markets'. *Global Crime* 14(2–3): 155–174.

Holt, T.J., and K.R. Blevins. 2007. 'Examining Sex Work from the Client's Perspective: Assessing Johns Using On-Line Data'. *Deviant Behavior* 28(4): 333–354.
Holt, T.J., K.R. Blevins and J.B. Kuhns. 2008. 'Examining the Displacement Practices of Johns with On-Line Data'. *Journal of Criminal Justice* 36(6): 522–528.
Holt, T.J., K.R. Blevins, and J.B. Kuhns. 2014. 'Examining Diffusion and Arrest Avoidance Practices among Johns'. *Crime & Delinquency* 60(2): 261-283.
Holt, T.J., O. Smirnova and Y.-T. Chua. 2016. *Data Thieves in Action: Examining the International Market for Stolen Personal Information*. New York: Springer.
Honick, R. 2005. *Software Piracy Exposed*. Sebastopol, CA: Syngress.
Hutchings, A. and T.J. Holt. 2014. 'A Crime Script Analysis of the Online Stolen Data Market'. *British Journal of Criminology* 55(3): 596–614.
Import.io. 2017. Import.io. https://www.import.io/.
Internet Archive. 2017. Wayback Machine. https://archive.org/web/.
James, L. 2005. *Phishing Exposed*. Sebastopol, CA: Syngress.
Kozinets, R.V. 2010. *Netnography: Doing Ethnographic Research Online*. London: Sage Publications.
Krebs, B. 2016. 'Mir Islam—the Guy the Govt Says Swatted My Home—to Be Sentenced June 22'. https://krebsonsecurity.com/2016/06/mir-islam-the-guy-the-govt-says-swatted-my-home-to-be-sentenced-june-22/.
Kruithof, K., J. Aldridge, D. Décary-Hétu, M. Sim, E. Dujso and S. Hoorens. 2016. *Internet-Facilitated Drugs Trade: An Analysis of the Size, Scope and the Role of the Netherlands*. Cambridge: Rand Corporation.
Leong, N., and J. Morando. 2015. 'Communication in Cyberspace'. *North Carolina Law Review* 94(1): 105–159.
MacDonald, Z. 2002. 'Official Crime Statistics: Their Use and Interpretation'. *The Economic Journal* 112(477): F85–-F106.
Martin, J. 2014. 'Lost on the Silk Road: Online Drug Distribution and the ‘"Cryptomarket."’ *Criminology and Criminal Justice* 14(3): 351–367.
McCarty, B. 2003. 'Automated Identity Theft'. *IEEE Security & Privacy* 99(5): 89–92.
Meijerink, T.J. 2013. 'Carding: Crime Prevention Analysis'. PhD. diss., University of Twente, The Netherlands.
Meyer, G.R. 1989. 'The Social Organization of the Computer Underground'. MA thesis, Northern Illinois University, DeKalb..
Mosendz, P. 2017. 'Middle-Aged Americans Beat Millennials in Time Spent on Social Media'. Bloomberg. https://www.bloomberg.com/news/articles/2017-01-25/middle-aged-americans-beat-millennials-in-time-spent-on-social-media.
Motoyama, M., D. McCoy, K. Levchenko, S. Savage and G.M. Voelker. 2011. 'An Analysis of Underground Forums'. Paper presented at the Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement, Berlin.
Munksgaard, R., J. Demant and G. Branwen. 2016. 'A Replication and Methodological Critique of the Study “Evaluating Drug Trafficking on the Tor Network”'. *International Journal of Drug Policy* 35: 92–96.
Nagpal, C., K. Miller, B. Boeking and A. Dubrawski. 2016. 'An EntityResolution Approach to Isolate Instances of Human Trafficking Online'. *arXiv preprint arXiv*:1509.06659.

Parks, M.R., and K. Floyd. 1996. 'Making Friends in Cyberspace'. *Journal of Communication* 46(1): 80–97.
Peretti, K.K. 2008. 'Data Breaches: What the Underground World of Carding Reveals'. Santa Clara High Technology Law Journal 25(2): 375–413.
Petsas, T., G. Tsirantonakis, E. Athanasopoulos and S. Ioannidis. 2015. 'Two-Factor Authentication: Is the World Ready?: Quantifying 2FA Adoption'. Paper presented at the Proceedings of the Eighth European Workshop on System Security, Bordeaux.
Poulsen, K. 2012. *Kingpin: How One Hacker Took Over the Billion-Dollar Cybercrime Underground*. New York: Random House Digital, Inc.
Reid, E. 1991. 'Electropolis: Communication and Community on Internet Relay Chat.' PhD diss., University of Melbourne.
Salus, P.H., and G. Vinton. 1995. *Casting the Net: From ARPANET to Internet and Beyond*. Boston: Addison-Wesley Longman Publishing Co., Inc.
Scraping Hub. 2017. Scraping Hub. https://scrapinghub.com/
Soska, K., and N. Christin. 2015. 'Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem'. Paper presented at the USENIX Security Conference, Washington, DC.
Soudijn, M.R., and B.C.T. Zegers. 2012. 'Cybercrime and Virtual Offender Convergence Settings'. *Trends in Organized Crime* 15(2–3): 111–129.
Speedtest. 2016. Speedtest® Market Report. http://www.speedtest.net/reports/united-states/
Tickle, K. 1994. 'The Vicarious Liability of Electronic Bulletin Board Operators for the Copyright Infringement Occuring on Their Bulletin Boards'. *Iowa Law Review* 80(2): 391–418.
Turban, E., J.K. Lee, D. King, T.P. Liang and D. Turban. 2009. *Electronic Commerce 2010*. Upper Saddle River, NJ: Prentice Hall Press.
Van Buskirk, J., R. Bruno, T. Dobbins, C. Breen, L. Burns, S. Naicker and A. Roxburgh. 2017. 'The Recovery of Online Drug Markets Following Law Enforcement and Other Disruptions'. *Drug and Alcohol Dependence* 173: 159–162.
Web Scraper. 2017. Web Scraper. http://webscraper.io/.
Westlake, B., M. Bouchard and R. Frank. 2012. 'Comparing Methods for Detecting Child Exploitation Content Online'. Paper presented at the Intelligence and Security Informatics Conference (EISIC), Odense, Denmark.
Yip, M., C. Webber and N. Shadbolt. 2013. 'Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing'. *Policing and Society* 23(4): 516–539.

[1] Some software scrape the content at the same time as it is downloaded. Whether the scraping work happens later or not does not impact the speed of the data collection.