"Big data and big crimes are powered by cloud technologies and the Internet of Things . . ."

How Big Data Feeds Big Crime

DAVID S. WALL

"Big data is the new oil," said Clive Humby, the British mathematician and marketing expert, back in 2006. But data, like oil, cannot be used unless it is refined into valuable products that create profit. Big data—the transactional and content-viewing data generated by our interactions with the Internet—must be analyzed, or refined, and made into information products. Only then does it acquire a value that others are prepared to pay for.

Big data helps businesses predict consumer behavior. You have probably noticed how curiously appropriate many of the suggestions from your favorite online vendors are. They analyze data from your past purchases, your movements, and your entertainment likes and dislikes in order to calculate what you would like to buy, eat, listen to, or watch. Big data analytics are also used for many other functions, such as weather forecasting, traffic bulletins, and some more contentious tasks like allocating policing resources or informing judges and parole boards about an individual's likelihood of recidivism.

Big data is a very disruptive phenomenon. It has brought a range of exciting new tools that offer great potential for identifying new truths about social and physical phenomena that were previously impossible to research on such a large scale. But it has become a generic buzzword for a disparate range of analytic technologies based on algorithms that are supposed to somehow accurately predict the future. There is greater scope for disruption when the analytics shift from anticipating possible events—ascribing probabilities on the basis of what has happened previously—to making predictions about when a specific event will occur or what specific individuals or groups will do in the future.

DAVID S. WALL *is a professor of criminology at the Centre for Criminal Justice Studies in the University of Leeds' School of Law.*

The commercial realization of value in these products has stimulated heavy demand for data—any data—and the creation and sale of tools to analyze it. Demand far outstrips supply, partly because the genuine value of big data products has been exaggerated by the claims of some analytic marketing campaigns.

This demand has also transformed some aspects of criminal behavior. Most of the illegal products and tools for big data analysis that help criminals can be obtained from illicit online markets such as AlphaBay, Silk Road, and Dream Market, to name a few. These markets are regularly taken down by law enforcement, but just as regularly reappear in a different format. The dark web is where the illegal markets and communications forums are found. It is part of the deep web, which underlies the visible Internet and is beyond the reach of search engines such as Google. The deep web is where the mechanics of social media and other communications operate. It can be accessed by tools such as the Tor network, which uses encryption to preserve users' anonymity.

Lucrative illicit markets for data have facilitated a range of what I call big crimes, even though the law relating to some of them is not yet clearly defined. These are largely "upstream" crimes such as data breaches, distributed denial of service attacks (DDoS), and mass spam attacks. Upstream crimes, usually committed against businesses, are damaging in their own right but also provide the information resources or capability for further crimes.

"Downstream" crimes take place when the stolen data is sold to unscrupulous types who may try to use it to exploit or extort either the individuals whose personal information has been compromised or the owner of the database, whether it is a business or another type of organization. This can lead to the disruption of services and businesses,

frauds, scams, hate speech, political interference, and so on. Even cyberterrorism and cyber warfare can flow from upstream data crimes.

Upstream crime ultimately creates losses and impacts on a scale larger and more threatening to individuals, businesses, and national infrastructure than any cybercrimes ever experienced before. It presents immense challenges for law enforcement agencies and society in general, not least because upstream and downstream crimes are often committed by entirely different actors against different groups of victims.

These different sets of actors and crimes can greatly complicate and frustrate the work of reporters, investigators, and prosecutors. Yet they are rarely the sole subject of police or criminological attention, which tends to mostly focus on downstream cybercrimes. Focusing more attention on upstream crimes could prevent or mitigate their downstream effects by denying criminals the information resources needed to commit other crimes.

Crime in the Cloud

Big data and big crimes are powered by cloud technologies and the Internet of Things, which amplify the distinctive qualities of the digital and networked systems that have transported data across the information superhighway—the Internet—for over a quarter of a century. In recent years, cloud technologies have rapidly expanded the capacity of the Internet, allowing it to work faster and cheaper. Cloud technologies are integral to most Internet services today. They allow data to be stored and accessed via the Internet instead of on the user's own hard drive.

By increasing computing capacity and power at a relatively low cost, cloud technologies act as a force multiplier. Evidence of this effect can be shown in terms of the exponential growth in computing jobs—or by looking at the increase in revenue recorded by the largest cloud platform, Amazon Web Services, from $1.05 billion in the first quarter of 2014 to $3.67 billion in the first quarter of 2017.

Such growth in volume and value makes big data attractive to investors and also to criminals. Not only do the latter recognize the advantages that big data brings with it, but they can also use it to reduce their personal risk. Internet technologies originally transformed criminals' options by enabling them (in theory) to replace one high-risk $50 million robbery with 50 million one-dollar robberies that could be committed in relative safety. The new force multiplier of cloud technologies now enables them to commit 50 billion ten-cent robberies with even less relative risk for a higher gain.

The emergence of the Internet of Things has further expanded the data flow by increasing the number and variety of devices gathering information. Your smartphone, your daily step-counter, even your car, household goods, and house itself are all generating data on your movements and decisions and communicating this data back to their motherships.

Together these technologies gather, store, and process big data, creating entirely new markets. This gives rise to new criminal opportunities for the misuse of that data in the form of data breaches (or data theft), large DDoS attacks, and mass spam attacks that send virus-bearing emails downstream in an attempt to compromise victims' computer systems. These three specific cyber-dependent big crimes each rely heavily on cloud technologies, and they each use big data in quite different ways.

Rampant Breaches

Equifax, Uber, Target, Yahoo, and LinkedIn are only a few of many major brand names that have been the victims of data breaches in recent years. Add to these a range of public sector and government agencies that have also been breached and the list keeps growing. At the heart of these crimes is the theft of big data. (Of course, the data is not stolen, strictly speaking; during a data breach, the data is copied and usually remains on the original system.)

Much of the data lost in such breaches is probably fairly useless to criminals seeking to exploit the primary owner. This is because email addresses and passwords tend to change often due to (still arguably weak) security practices. However, in addition to the stolen data set being used against its primary owner (a large company such as Equifax, or the other corporate victims), it can also be combined with other stolen data to be used with greater effect against the individual customers of that company.

To give some idea of the scale of these incidents, the Identity Theft Resource Center and CyberScout reported that the number of data breaches increased in 2017 by almost a third from the previous year, rising from 1,000 in 2016 to around 1,300 in 2017—and was up from 200 in 2005.

Since 2005, information technology firm QWERTY Concepts reported, there have been 7,700 breaches that exposed almost a billion personal records. The number of different sectors being hit is expanding. The business and medical sectors appear to have become particularly vulnerable to attack in recent years.

One interesting trend is that the average size of single data breaches may be decreasing slightly in terms of the cost of each compromised customer record. The 2017 Ponemon Institute report on data breaches found that their cost had decreased slightly to $141 per record from $158 in 2016 and $154 in 2015, compared with a rise from $202 to $214 between 2008 and 2010. This decrease is attributed to a strong dollar and also to the expanding presence of incident response teams either in-house or on contract.

However, organizations' collective loss from reported data breaches in the United States (and elsewhere) has increased dramatically because of the sheer volume of breaches experienced today. And not all data breaches are reported immediately, or at all: in November 2017, the ride-hailing company Uber disclosed that it had kept secret a breach of some 57 million records of drivers and riders for more than a year.

If businesses can use big data, so can tech-savvy criminals.

If businesses can use big data, so can tech-savvy criminals—especially financial information, as the 2017 Equifax breach indicates. Equifax, a US credit-monitoring agency, admitted that the personal data of 145 million US consumers (and 400,000 UK consumers) had been accessed or stolen in a massive hack in May 2017, following a major data breach two months earlier. The implications of these breaches are still emerging, but some commentators have suggested that they could ultimately prove so serious that credit systems may have to be overhauled and strict new laws passed to better protect customers.

Big data breaches cause serious problems for the organizations from which the data is stolen, in terms of both business losses and reputational damage, not to mention the costs of restoring systems and of installing and maintaining more sophisticated security measures. They also create subsequent problems for the individuals or organizations downstream whose data is stolen, especially when that data is augmented with data from other breaches.

Hypothetically, data from a basic breach of non-financial information can be linked by an email address to, say, financial information from another breach, and then linked, perhaps by Social Security number, to other breaches relating to health or taxes. The resulting big crime data set increases dramatically in value because this composite personal profile can be used much more effectively—with more certainty and possibly less risk. So it can garner a much higher price when sold to criminals downstream.

YOUR MONEY OR YOUR DATA

Big data crimes are still in their early growth phase, but they continue to evolve. In addition to the criminal resale of illegally obtained data for financial gain, consider these five other types of big crime linked to data breaches, each with its own motivations and methods.

First, there are whistleblowing websites such as WikiLeaks that publish leaked data. The motivation behind these leaks is usually to publish data in the public interest, so the hackers would not personally regard it as stolen—but prosecutors often do.

Second are ransom hacks: hackers demand a ransom from businesses to delete the stolen data and threaten that if they are not paid, they will sell it to a competitor—or publish it, as was the case in the hack of Ashley Madison, a dating agency for people seeking extramarital affairs. The problem with ransoms is that the victim has to trust the hackers to delete the data.

Third, ransomware that encrypts business data to render it inaccessible is becoming increasingly prevalent. It is usually activated when targeted spam emails sent out by mass spam attacks trick recipients into opening an infected attachment or website. Once they gain access to a business's computer system in this way, the hackers encrypt the data and demand a ransom to be paid in bitcoin or another cryptocurrency to unlock it. The problem here is also one of trust: the payment, which is usually discouraged by law enforcement, does not always result in a successful recovery of the data. What the big players tend to do on such occasions is to send victims a token "proof of life" (usually a decrypted file) in order to build trust.

Of course, if more business continuity plans contained a backup and rebooting strategy for key data, that would drastically reduce the need

to make ransom payments in the first place. Many companies and organizations, especially small to medium enterprises, believed until recently that they were immune to such attacks, assuming that they were not a big enough target for anyone to bother hacking them.

Fourth, with data theft there is always the problem of "sleeper fraud": data on individuals in either a personal or a business capacity is held in reserve until it can be combined with other data and eventually used to defraud them or disrupt business activities. This also increases the problem of repeat victimization.

A fifth scenario is the use of composite big data sets to build up a detailed information portfolio about businesses and their activities in order to commit industrial espionage. This may involve the application of data analytics tools to accumulated legal and illict data sets, which can unfairly identify revealing trends.

OVERWHELMING BOMBARDMENT

DDoS (distributed denial of service) attacks are a different type of big data crime. The perpetrators, for a variety of reasons, bombard online services with massive floods of data to overwhelm their access portals and thereby deny access to legitimate users. The growing impact and availability of cloud technologies, combined with the weaponization of Internet of Things devices via botnets, are increasing the size, scope, and complexity of DDOS attacks. Botnets are sets of "zombie" computers that have been infected by remote administration tools (malware) and can be controlled remotely by a third party (the "bot herder") without the owners' consent or knowledge to send out spams or commit DDoS attacks.

Big data is attractive to investors and also to criminals.

The cybersecurity firm Arbor Networks notes in a 2017 report that cloud services themselves are increasingly becoming the target of DDoS attacks, rising from 19 percent of the total in 2014 to 33 percent in 2016. Arbor also observed 558 attacks in 2016 with a speed of over 100 gigabytes per second, compared with 223 in 2015. These figures indicate a dramatic increase in the power of DDoS attacks. Ten years or so ago, attack speeds were more likely to be in the high megabytes per second, as much as 1,000 times slower than today.

DDoS attacks against companies and organizations can be particularly devastating, since they disrupt the continuity of business. They can also be easy to pull off, since they do not require attackers to have special knowledge. Anyone can simply subscribe to a site located on the dark web that offers a range of attack plans.

The motivations behind such attacks tend to be much broader and more mixed than for data breaches. Very often an attack is intended to facilitate some other form of cybercrime. Typically a DDoS attack is used either to weaken a security system before removing data, or to distract security personnel while another means of entry to the system is being used.

Alternatively, many DDoS attacks are carried out to exact some form of moral or political revenge, as was the case in attacks by hacker groups such as Anonymous against a number of organizations. DDoS attacks may be religiously (or emotionally) motivated—the attackers might be jihadists or just people disgruntled with a particular organization or individuals within it.

Another important motivation in such attacks, as with many cybercrimes, is the intellectual challenge involved. A novice hacker may cut his or her teeth on DDoS attacks, since they are known to be a relatively easy kind of cybercrime. Sometimes the motivation is just a desire to impress friends. Very often the hackers are still in a computer-game mindset and do not recognize the consequences of their actions. Alternatively, they may seek to establish a personal reputation within a hacker forum.

The tipping point into cybercrime may be a combination of such motivations. Financial gain is clearly a motive in some DDoS attacks but not all, and often it seems to come last on the list.

INFECTIOUS SPAM

As with data breaches and DDoS attacks, mass spam attacks are a type of big crime that sends out heavy loads of data across systems. Along with this upstream effect, it also has a downstream impact: the delivery of poisoned payloads designed to infect, extort, deceive, defraud, or steal from recipients. Mass spam attacks have increased exponentially since the advent of botnets in 2003 and subsequently with cloud-based technologies and the Internet of Things. Instead of 50 million emails per hour, 50 billion emails can now be sent every 20 minutes. The "hit rate" of spam emails

is relatively small, but the sheer volume of spam increases the overall number of hits.

While more sophisticated anti-spamming technology does tend to filter out most spam, many spam emails still hit the mark because of the increased ingenuity of spammers' tricks to avoid detection and fool the recipient into responding. Also, the bot herders who run the botnets that send out mass spams are constantly looking for more vulnerable computers to infect. One method for doing this is sending spam to new email addresses collected from data breaches. This enables them to increase the size of their herds of zombie computers and their ability to send out even more spams, thus increasing the value of the botnet.

The main threats from spamming are therefore upstream. Spamming clogs up bandwidth, but also encourages the market for stolen data (email addresses). More importantly, downstream criminals pay upstream spammers to send out their spams. Sometimes this is done through a direct relationship, but it is increasingly managed through cybercrime service sites that are available on the dark web.

The motivation of both upstream and downstream spammers is mainly financial. While the former exist to send out spams, the latter seek through their various payloads to victimize recipients in different ways. The emails sent out in mass spam attacks mainly try to manipulate the recipient into buying something, or divulging something such as personal information, or clicking on an infected link.

HUNTING THE HUNTERS

Big crimes are all cybercrimes: criminal behavior transformed by networked and digital Internet technologies. More specifically, as a genus of cybercrime, big crimes—data breaches, DDoS attacks, and mass spam attacks—are all cyber-dependent crimes: they are the product of Internet technologies, so if you were to take away the Internet they would disappear.

This is in contrast to cyber-assisted crimes—for example, organizing a drug deal or planning how to murder someone—which would still take place if the Internet were removed because the perpetrators would use other means to organize or work out how to commit the crimes. Cyber-dependent crimes are also different from cyber-enabled crimes (such as frauds), which would still occur, but on a more local scale, without the global reach of the Internet.

Different categories of big data crime can be distinguished in terms of the modus operandi: crime against the data (data breach), crime using the data (DDoS), and crime in the data (information that can be used against the owner). By differentiating the level of technological transformation and also the modus operandi involved in these types of crimes, we can identify the level and type of technology needed to counter them, as well as the applicable body of law.

As the 2002 movie *Minority Report* presciently suggested, big data analytics work very well when used to anticipate possible crimes, but largely fail to predict who the actual perpetrators and victims will be (though the predictors in the film were psychics known as precogs, not computers). Current computers create too many false positives to identify evidential proof beyond a reasonable doubt. What may be possible, however, is that the algorithmic ideas behind "anticipatory" analytics (not predictions) could be used to pinpoint when upstream crimes like data breaches, DDOS attacks, and mass spams take place.

This is something that my computer science colleagues in an ongoing research project are working on. They are seeking to use machine learning and artificial intelligence to find algorithms in the data that would detect such upstream attacks as they occur. The idea is that this will allow practitioners to identify patterns of data crime, which could lead law enforcement to the perpetrators.

Such an approach, if successful, raises a number of ethical issues, including the potential for reinforcing stereotypes of specific social groups as being overly criminal, as well as questions about admissibility of evidence in court. It will require a robust legal framework and practices such as transparency that would make it clear how the technology works.

DIGITAL DELINQUENTS

Big data has created opportunities for big crimes. But who are the perpetrators? Many of these crimes appear to be committed by teenagers. The UK's National Crime Agency reported in 2017 that the average age of those arrested for such offenses is 17 and under, with very few if any previous convictions in this or any other field. This profile is far from that of the street criminals normally found in the justice system, who commit increasingly serious offenses and are subjected to progressively harsher correctional regimes.

The big crime offenders belong to what I call the "flat food generation." This is a term borrowed from Douglas Coupland's 1995 novel *Microserfs*, which describes a generation of young people obsessed with computing. They lock themselves in their rooms and are saved from starvation only by being fed "flat foods" (such as Pop-Tarts), which their parents or friends slip under their doors.

Today's flat food generation is often seduced by cybercrime. They drift from playing video games into cheating by disabling their friends' computers. They go on to mix with like-minded people in online hacking chat rooms, learning the tricks of the trade, and eventually committing serious crimes such as DDoS attacks, which can result in substantial prison terms.

These are often very young people who are neither socially nor psychologically prepared for the mainstream criminal justice system. What we do not want is to see them being jailed and then falling under the wing of hardened organized criminals, who could call in favors once they have been released from prison. That could unleash a serious cybercrime wave.

Until recently, traditional organized crime groups tended to keep away from most online activities. Big crime is generating the financial incentives to build a new online mafia that would operate along the lines of the offline mafia by putting criminals under its protection and investing its profits to ultimately create wealth, influence, and political power.

The solution would involve combining reduced sentences for these youths—some punishment is needed to deter recidivism—with programs that redirect their skills. Alongside a real-time detection system and an improved legal framework to deal with data theft, protect individuals' privacy, and shape such redirection programs, there is a need for a serious preventive effort. This would involve, among other things, training school liaison police officers to address any strange or dangerous behavior early on. We need to turn these young people from black-hat hackers into white-hat hackers. This requires large-scale prevention programs with constructive alternatives to the criminal justice system that could eventually put those hacking skills to good use.

Big cybercrime is here to stay because we are in the age of big data. This is a pill that cannot be sweetened. Protective measures such as data backups, personal recovery tactics, and business continuity strategies can go a long way toward mitigating the damage done by increasingly common attacks. But we need a combined approach to big crime that integrates technological defenses with social and educational reforms, as well as improved legal procedures. Such an approach must also clearly define which government agencies are responsible for tackling the threat of big crime, recognizing that it has the potential to severely disrupt our economies and society in general. ■