# Artificial Intelligence Enabled Cyber Fraud: *A Detailed Look into Payment Diversion Fraud and Ransomware*

*Alana Maurushat*[*], *Abubakar Bello*[**] *and Braxton Bragg*[***]

ABSTRACT *Cyber fraud is rampant. The recent Covid 19 pandemic is a good example of the same. Domain Tools in April 2020 identified over 65,000 websites have been identified as fraud scams related to Covid-19. Organisations have lost billions of money in online scams, and in particular with payment diversion fraud ('*PDF*') and ransomware. PDF is a type of cyber-attack where an entity is tricked into making a direct payment from its account to a false supplier/entity often using real-time payment methods. Ransomware is a type of malicious software that prevents users from accessing their system or personal files usually by locking them through encryption, and demands ransom payment in order to regain access. Based on the professional experience of the authors, coupled with current literature, there is a growing trend of automation, with the use of machine-learning and artificial intelligence. This article discusses PDF and ransomware in the context of mechanics and emerging trends for systematic attacks and response by private industry. These case studies illustrate the limited role that the law plays in the investigation and response to cyber fraud.*

[*] Alana Maurushat is Professor of Cyber security and Behaviour at Western Sydney University and a Board Director with the cybercrime investigations firms IFW Global.

[**] Dr. Abubakar Bello is Lecturer in Cyber security and Behaviour with a strong industry background in information security, risk management and digital forensics. He is also an expert in Nigerian cybercrime. Both are researchers with the Socially Engineered Payment Diversion Fraud (SocEngPDF) project.

[***] Braxton Bragg is a Research Assistant with SocEngPDF; he is a US-licensed attorney completing a Masters in Cyber security and previously held legal and accounting roles with private US firms. Thank you to student intern Kevin Tang for his research.

## I. Introduction

Artificial Intelligence Enabled Cyber Fraud encompasses a wide range of traditional online fraud using new tools to automate aspects of the process. This article focuses on two primary online frauds: payment diversion and ransomware. We will go through each type of fraud, first explaining the concept, providing a case study, and then addressing threat vectors – commons ways in which the fraud is committed from financial data to accounting practices, to network intrusion to social engineering aspects. Automated AI aspects are explored within the threat vectors. The case studies are based on real cyber frauds, but the names, personal information, and case specific details have been generalised to protect the identity of the parties involved. This is especially important as these types of audits and investigations are often done over many years, and can involve civil litigation and criminal charges.

As many of these investigation case studies are based on the experiences of the researchers, we need to make clear the capacity and background of the researchers in question. Alana Maurushat is the Professor of Cybersecurity and Behaviour at Western Sydney University, as well as Board Director of the internally renowned cybercrime investigation company, IFW Global. IFW Global is renowned for taking down organised cybercriminal syndicates and recovering funds for individuals and organisations. This article is not underpinned by theory; it is based on first-hand experience of the authors in their roles as researchers, and expert consultants. Where possible, we have cited news articles, television programs, and white papers produced by the industry partners and organisations where the researchers work as expert consultants. Dr. Abubakar Bello is an expert in digital forensics and incident response. He works as a consultant to major companies and government agents. Braxton Bragg is a tutor with the Cyber security and Behaviour program at Western Sydney University, and is senior cyber security consultant with Gridware. He comes from a background in forensic accounting, incident response as well as cyber security compliance.

Maurushat and Bello are currently finalising two funded research projects: Socially Engineered Payment Diversion Fraud, and Ransomware. We have done qualitative interviews with over 30 organisations, and online quantitative analysis from an additional 150 organisations via an online survey – all of which have had recent experiences countering funds lost due to payment diversion fraud or ransomware. The findings of these research projects are not included in this article as the research hasn't been finalised and the work is currently under peer review. Some of the case studies in this article, however, have been built around the work done within these research projects, as well as from first-hand experience of the researchers in their capacity as investigators and consultancy work with industry.

There is a paucity of peer reviewed research articles globally that provide first-hand experiences of some of the problems that law enforcement faces when dealing with certain types of cybercrime, and even less for AI enabled cybercrime. There is a plethora of research on policing and legal approaches, for example, to online child pornography, and online copyright. There is significantly less research on policing and legal responses to online fraud and cyber attacks. Many of these wider cybercrime articles focus on fraud typographies,[1] reviewing of data found within media and

[1] Rodger Jamieson and others, 'Addressing Identity Crime in Crime Management Information Systems: Definitions, Classifications and Empirics' (2012) 28(4) Computer Law & Security Review 381.

blogs,[2] or they provide a statistical and economic analysis.[3] There are a number of reasons for speculation as to why this is the case – it could be that researchers are not as interested in online fraud as other forms of cybercrime, or, as per the experience of the authors, the police have limited budgets and are predominantly focused on crimes where the elements happen within a set jurisdiction. If law enforcement has a limited budget for expensive complex cases, they will focus on the areas where the harms are perceived as the greatest – online child abuse and elements related to national security. Cyber attacks such as payment diversion fraud and ransomware require rapid investigatory response if funds are to be recovered. Police are not set up to deal with these types of investigations. For this reason, private firms are called in to do most of the investigatory work alongside law enforcement. This private work is carried out by cyber security and cybercrime experts typically found in consultancy firms such as Price Waterhouse Cooper, EY, large major law firms, as well as smaller boutique of asset recovery firms. Private firms play an essential and dominant role in countering many types of cybercrime.[4]

The article first addresses Payment Diversion Fraud in Part 2, followed by Part 3 which addresses Ransomware. The next section, Part 4, outlines the criminal legal framework for these areas with fraud being the main area of relevant law. It outlines appropriate laws that deal with these types of fraud by looking at the Convention on Cybercrime, then at relevant provisions in Australian, Canadian and Indian criminal codes. Part 5 addresses why online fraud has one of the highest payouts of cybercrime with the least risk, and examines why law enforcement are ineffective at investigating organised online fraud, prosecuting offenders and recovering fraudulent funds to victims. This part has been written explicitly to capture the sentiments expressed in qualitative interviews from two funded research grants: Socially Engineered Payment Diversion Fraud, and Ransomware. Part 6 looks at ways to reform online fraud investigations. The final section, Part 7, offers concluding remarks. Annex at the end of this article contains a list of essential terms with their definitions.

[2] *See* for example, Roderic Broadhurst and others, 'Crime in Cyberspace: Offenders and the Role of Organized Crime Group' (2013) <https://ssrn.com/abstract=2211842> accessed 6 May 2020.

[3] *See* for example, Lina Fernandes, 'Fraud in Electronic Payment Transactions: Threats and Countermeasures' (2013) 2(3) Asia Pacific Journal of Marketing & Management Review 23.

[4] *See* Alana Maurushat and Hadeel Al-Alosi, 'Policing Cybercrime – An Inside Look at Private and Public Cybercrime Investigations' in Philip Birch (ed), *Australian Policing: Critical Issues in 21st Century Police Practice* (Routledge, Forthcoming 2020).

# II. Payment Diversion Fraud

## A. What is Payment Diversion Fraud?

Payment Diversion Fraud is a type of cyber-attack where an entity is tricked into making a direct payment from its account to a false supplier/entity often using real-time payment methods.[5]

Payment Diversion Fraud has been around for several decades but didn't emerge as its current form until recently. Previously one would have described PDF as a man-in-the-middle-attack but it didn't connote what happened after the attack, namely a fraudulent act.[6] Other terms associated with PDF are supply chain fraud,[7] mandated fraud[8] and business email compromise.[9]

To date there is limited research and analysis on PDF in the public domain.[10] PDF is related to another concept in what is referred to as 'compromised' or 'poisoned' supply chains whereby at any point in the supply chain for a product development or service provided, there are multiple vulnerabilities that can be compromised.

Payment Diversion Fraud has a great economic impact on organisations, with U.K. law enforcement characterising it as the most harmful reported fraud with greater economic impact than Brexit[11] and the U.S. FBI describing it as one of the costliest forms of cyber-enabled fraud affecting U.S. companies. Earlier PDF used phone calls and phishing emails. More recently media have reported fraudsters gaining unauthorised access to an entity's network/phone/IoT—monitoring the network to observe business and

[5] Ken Gamble, 'Payment Diversion Fraud – A disturbing new hacking trend hitting corporate Australia' (*Akolade*, 13 February 2018) <http://akolade-blog.blogspot.com/2018/02/payment-diversion-fraud-disturbing-new.html> accessed 6 May 2020.

[6] Twenty Essex, "Man-in-the-middle" fraud: How to prevent it, who is at risk, and what to do when it all goes wrong' (*Lexology*, 25 April 2017) <https://www.lexology.com/library/detail.aspx?g=6c8cce34-0bfe-4aa6-86fc-edfa88e7c473> accessed 20 March 2019.

[7] James L. Patterson, Kimberly N., Goodwin, and Jennifer L. McGarry, 'Understanding and Mitigating Supply Chain Fraud' (2018) 12(1) Journal of Marketing Development and Competitiveness 70.

[8] Paul Dean and Rory Grout, 'Something rotten in the state of shipping: What you need to know about Mandate Fraud and the fraudulent redirecting of payments' (*HFW*, October 2017) <http://www.hfw.com/Something-rotten-in-the-state-of-shipping-what-you-need-to-know-about-Mandate-Fraud-and-the-fraudulent-redirecting-of-payments-October-2017> accessed 20 March 2019.

[9] Federal Bureau of Investigation, *2017 Internet Crime Report* (2017).

[10] Steven Powell, 'Critical Measures to Protect Against Rocketing EFT Fraud Risk Management' (2009) 9(11) Without Prejudice 48.

[11] Mara Stein, 'UK Companies Plagued by Payment Diversion Fraud' (*The Wall Street Journal Blog*, 6 October 2017) <https://blogs.wsj.com/riskandcompliance/2017/10/06/u-k-companies-plagued-by-payment-diversion-fraud/> accessed 6 May 2020.

cultural patterns of the organisation before sending out what appears to be a legitimate request from a CEO/finance department/supplier in the form of an email, text, or similar requesting payment.

Payment Diversion Fraud is being committed by a range of criminals located around the world. It is not jurisdiction-specific, though many recent cases derived from our current research grant on PDF have involved payment being made to accounts in Hong Kong, Ukraine, South Africa, Ghana, Nigeria, India, and Brazil. Two anonymised incidences with company IFW Global involved physically tracing payments back to the source located in Nigeria and India (cyber and on the ground surveillance). Our mere cyber investigations of following money trails led to Hong Kong, Ukraine, South Africa, Ghana, and Brazil.

The methods used to enable a PDF attack tracks with a fairly-standardized process used in other types of cyber-attacks. The process progresses through the following sequence: conducting initial reconnaissance, conducting the initial compromise, establishing a foothold in the system, escalating privileges, conducting internal reconnaissance, moving laterally in the system, maintaining a presence in the system, and completing the mission. The main variant is the extent to which automation and artificial intelligence are now part of process.

### i. Initial Reconnaissance of Organisation Through Public Information

In this phase, the threat actor will engage in initial reconnaissance of possible targets. Many times, this reconnaissance is performed using open source intelligence (**'OSINT'**). OSINT uses many sources generally available to the public on the internet[12] (eg Google searches, Linked-In searches, social media sources, news articles, and conference websites). The data gleaned from this intelligence gives the threat actor precursory knowledge of potential victim organisations, before ever deciding which organisations to attack. The gathering of information available to threat actors through this method is almost impossible to stop, as customers, vendors, and partners need to have a method of gaining information about organisations with whom they want to engage.

Sometimes this process is automated in the same way that crawlers used by Google, and other search engines, retrieve results for search queries. The attacker does not have to necessarily invest large amount of time to discover

[12] MITRE Corporation, 'Acquire OSINT data sets and information' (*MITRE ATT&CK*, 14 December 2017) <https://attack.mitre.org/techniques/T1247> accessed 27 March 2019.

potential targets. Scripts using common web scraping computer programs can be combined with industry-specific lists of organisations to compile OSINT for targeting would-be victim organisations. Web scraping is also known as web harvesting or web data extraction - web scraping means extracting data from websites in a usable and structured format which is done through a variety of web tools. Social network sites like LinkedIn make this type of open source intelligence gathering easy and provide a wealth of information. The outputs can then be combined with other OSINT data points such as expected travel plans (eg via conference schedules, or news articles) to put attackers on notice of potential targets. AI-based machine learning algorithms can also be used to piece together these data points faster and more effectively to help target more victim organisations.

A simpler method of initial reconnaissance can begin with a phone call to a member of the targeted organisation. Telephone conversations can be used for finding intelligence, gaining trust and learning the behavioural aspects of employees at a firm. And AI techniques can be used to help with this technique too. Some security researchers are already concerned about new technologies such as Google's deep neural network-based Duplex service that can be trained to interact over phone lines with humans without the humans ever knowing a computer is involved.[13] In traditional AI systems, machine learning uses computers to process and learn from data. With neural networks programs try to emulate how the brain processes information with an input layer, output layer, and multiple hidden layers that interact with one another simultaneously. With deep learning the computer trains itself to process and learn from data. Deep neural networks are a method of a computer training itself to process and learn from data mimicking processing of a human brain, and in the case of AI, by additionally mimicking human behaviour. A substantial fear is that threat actors could train these services to conduct intelligence in a highly automated process.

There are other times when no initial reconnaissance is done. In these cases, the process would begin with the phase of establishing access in an organization's network, but this tends to be more common with ransomware rather than payment diversion fraud.

[13] Yaniv Leviathan, 'Google Duplex: An AI System for Accomplishing Real World Tasks Over the Phone' (*Google AI Blog*, 8 May 2019) <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html> accessed 6 May 2020.

### ii. Accessing a Weak Point

In this phase, the threat actor will gain some type of initial access to a potential victim organization's systems. Methods of establishing a foothold can include standard phishing emails, spear-phishing emails, man-in-the-middle (MITM) attacks, watering hole attacks, password spraying, or drive-by downloads. The various methods used to gain initial access into an organization's systems are called threat vectors and are discussed below in section 2.3.

### iii. Accessing Escalated Privileges

Once initial access to the network is obtained, the next step in the cyber attack is to escalate privileges to allow movement through the network undetected. Privileged access, normally administrator-level, is needed because it allows the attackers to move freely within the environment and remove traces of ever being there. Sometimes rainbow tables and similar tools can help intruders steal credentials. On other occasions, attackers use threat vectors like spear-phishing emails from within the system to help them escalate privileges, usually allowing them to access any system on the network. Once the attackers gain elevated privileges, the network is effectively taken over and 'owned' by the intruders. This allows them to take on the next step of the attack by conducting internal reconnaissance.

### iv. Conducting Internal Reconnaissance of the Organisation's Network

Many times, attackers can be in a network for months conducting internal reconnaissance. Recent intelligence reports show that the average time before a network intruder is detected, called the dwell time, in the APAC region during 2018 is 204 days, down from 498 days in 2017.[14] During this phase, attackers are looking for vulnerabilities and examining accounting practices, calendars, company directories, invoice and payment protocols, the tone and rhetoric of emails, and other information that can help achieve the fraud.

The median dwell time is reducing quickly due to organisations gaining a better understanding of best-practices in mitigating their cyber security risk and using more advanced security systems, which many times use AI-based processes. However, while organisations are using AI to help prevent and

[14] Fireeye Mandiant, 'M-Trends Report 2020' (2020) <https://www.fireeye.com/current-threats/annual-threat-report/mtrends.html> accessed 6 May 2020.

detect attacks, attackers are also starting to use AI to make their attacks more effective. In 2017, cyber security firm Darktrace found that attackers were using AI, through machine learning algorithms, to observe average user behaviour in a client network in India.[15] The attackers were then able to use their AI software to mimic this average user behaviour, allowing them to stay undetected in the network for a longer period of time. This is just one example of AI being used to help attackers. Very recently, a report compiled by 26 authors from 14 institutions made predictions about the landscape of malicious use of AI over the next five years.[16] The report discusses scenarios including automation of vulnerability discovery and increased effectiveness of vulnerability exploitation. The report also discusses how the downward moving cost of AI will enable underfunded attackers to use advanced techniques. Overall, it paints a very bleak picture of how malicious use of AI could be employed by attackers in a variety of ways. Once attackers are in an organisation's systems, they normally take measures to ensure they can stay there.

### v. Sustaining a Presence

At this stage, although the attackers are in an organization's network with unrestricted access, they must take steps to ensure they are able to sustain a presence long enough to complete the fraud. To accomplish this, sometimes they install malicious programs like root kits and backdoors that allow them to return as frequently as they want, even if they are detected. On other occasions they create ghost users, fake employees with elevated access, and hide those users from real administrators. There are also a variety of other techniques used to ensure continued access. At this point, the original attack vector used to gain access is no longer necessary, and an organisation that remedies the original vulnerability is often left in a worse position than they started. The organisation thinks the intruders are gone and its risk has diminished, but in reality, the attackers can come and go as they please.

### vi. Figure out Precisely When and How to execute Payment Diversion Fraud

Using the information gleaned from the internal reconnaissance, the attackers will now create a game plan for conducting the fraud. They understand

[15] Steven Norton, 'Era of AI-Powered Cyberattacks Has Started' (*The Wall Street Journal Blog*, 15 November 2017) <https://blogs.wsj.com/cio/2017/11/15/artificial-intelligence-transforms-hacker-arsenal/> accessed 6 May 2020.

[16] Miles Brundage and others, 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation' (*Arvix*, 20 February 2018) <https://arxiv.org/pdf/1802.07228.pdf> accessed 6 May 2020.

how the invoice and payment process flows, whom, or what systems to communicate with, and how they will execute the fraud. But many times, they will conduct a test run before trying to 'swing for a six'.

### vii. Testing a Payment Diversion Fraud with a Low Sum

In many cases, organisations affected by PDF will be hit by multiple payment diversions. The first one or two diversions will be attempted for low value amounts that are unlikely to set off any alarms. In the test runs, and the final fraud payouts, email hijacking or spoofing is the primary method of conducting the fraud. In both cases the attacker will often wait for an opportune time (eg when an account signatory is unavailable due to travel, or a vendor payment is expected and there is a time crunch on receiving the payment).

In an email hijacking, the attackers use a signatory's actual email account to send a request for a payment to be made to an account controlled by the attacker. Sometimes this payment is an expected payment, such as a previously intended payment to an actual vendor. On other occasions the payment will be to a fake vendor setup in the organization's systems by the attacker. Still other times the payment will be a non-expected payment to a real vendor diverted to an attacker-controlled account.

In an email spoofing, the attacker will use a spoofed or look-alike email address to request a payment. A spoofed account can be easily created using very little technical knowledge to make a recipient think the email actually came from a signatories account, although the attacker never actually had control over the real account.[17] Though there is almost always some type of access that has been gained by the attacker to enable enough internal organizational information to plan the attack.

In a look-alike email, the attacker will create an email address on an attacker-controlled domain that seems, to the average user, to come from a signatory's actual address. For illustration: a real account would be named *bob_smith@organisation.com*, but the look-alike email would be *bob_smith@organisat*an*.com* (emphasis on the changed letter). In both types of attacks, the attackers will also likely create or spoof email addresses on both the sending and receiving side, so they are able to communicate to both sides of the transaction thus enabling them to continue to perpetuate the fraud.[18]

[17] Dylan Tweney, 'How to Fake an Email From Almost Anyone in Under 5 Minutes' (*Hakernoon*, 26 October 2017) <https://hackernoon.com/how-to-fake-an-email-from-almost-anyone-in-under-5-minutes-12169dd44a92> accessed 26 March 2019.

[18] Gamble (n 5).

For example, the attacker would be impersonating a vendor to get an organisation to make a payment, while at the same time impersonating the organisation so that the real vendor doesn't know the fraud is occurring either. Other types of diversion could include creating fake vendors and creating payment instructions for an EDI system to divert funds. There are also many other tactics used during PDFs. But as we said, this is likely just a test run.

### viii. Complete the Mission

Once the attackers have tested the plan for the fraud with a low amount that is not easily detected by normal accounting review procedures and is sure the fraud will work, they normally will try to make a large final payout. They will follow the same process as before, but this time, they will try to obtain an amount that will be material to the financial statements of the business, and will likely be discovered through normal account reconciliations. At this point, the attackers are finished with their activities and have likely left the network of the organisation for good. But organisations still need to ensure that a complete forensic analysis is conducted on their networks and end points to ensure that the attackers cannot return. Now we will explore some case studies to further describe real-life situations we have seen.

## B. Case Study

The facts of these cases, while resembling real investigations, have been altered to protect the affected parties involved in what may be ongoing disputes and investigations.

### i. Case A: Large International Company - Sport Equipment Retailer

Company X is an Asian based sport equipment retailer with annual turnover of USD 100 million. They are not, however, listed on the US stock exchange, and are not a publicly traded company.

Company X was notified in October 2017 by one of their suppliers, Supplier Y, that they had not received due payment of USD 2.1 million. Company X, however, claimed that they had paid Supplier Y in September 2017. Upon further investigation by both entities, it was discovered that Company X had been the victim of payment diversion fraud.

The CEO of Company X was flying to Malaysia on business in September 2017. An email generated from his iPhone 7 was sent to the company accounts receivable as he boarded Flight A. The email requested immediate payment of the attached invoice of USD 2.1 million as the CEO had 'forgotten to instruct payment' before he left the office for vacation. The invoice had been generated on Supplier Y letterhead with the details identical to previous invoices but for a change in Swift Code and banking details. The employee at accounts receivable read the email and made the payment. The payment of USD 2.1 million arrived in a bank account in Hong Kong registered to the name of a shell company (registered in Luxemburg) that was not affiliated with Supplier Y. From Luxemburg the money was sent to an additional three different shell companies with accounts located in various tax haven jurisdictions.[19]

An examination, audit, and investigation of the PDF revealed the following things. First, it was discovered that the data log files from the period of June 2015 up until May 2016 showed that Employee C had been compromised. Employee C had systems admin clearance and upon closer scrutiny appeared to have some slightly unusual activities for a period of close to a year. Employee C was also on the Company X's whitelist.[20] Employee C's accounts had been high-jacked. Slowly a very sinister pattern emerged. It was later revealed that Employee C's account had sent a phishing email to the CEO in December 2016 which resulted in the download of malicious software enabling the criminal(s) in question to install a rootkit onto the CEO's desktop computer. The iPhone 7 was also compromised but it remained unclear how this was achieved. This could have been done by the use of a known unpatched vulnerability as found on the dark net. A forensics examination of the phone did not reveal anything unusual though it was reported in September 2017 that Google successfully released a proof of concept attack against a Wi-Fi firmware vulnerability in Broadcom chips using a backdoor into the iPhone 7.[21]

Further scrutiny into payments made over a one-year period revealed that there were in fact three separate fraudulent payments made to third parties. The amounts started as nominal and involved fraudulent invoices from a range of what looked to be normal suppliers. In the first two instances, the email accounts from another senior employee were used. These appeared to be tests prior to the 'big heist' involving the USD 2.1 million using the CEO's email account and iPhone.

[19] Richard Murphy, 'World's Best Tax Havens' (*Forbes*, 6 July 2010) <https://www.forbes.com/2010/07/06/tax-havens-delaware-bermuda-markets-singapore-belgium.html#6a3819b825fc> accessed 25 March 2019.

[20] Whitelisting is the practice of explicitly allowing identified trusted entities access to a particular privilege, service, mobility, access or recognition.

[21] Michael Mimoso, 'Remote Wi-Fi Attack Backdoors iPhone 7' (*Threat Post*, 27 September 2017) <https://threatpost.com/remote-wi-fi-attack-backdoors-iphone-7/128163/> accessed 25 March 2019.

Upon further scrutiny of Company X's network it was later discovered that there was a dormant piece of malicious code that sent messages back to what appeared to be a range of IP addresses in what was suspected to be the various command and controls of a botnet. This suggested that first, the network had been compromised for approximately a year. Second, that firewalls, anti-virus and all other cyber security software were ineffective at detection. Third, that the silent, hidden surveillance aspect involved automation, and possibly elements of artificial intelligence. Last, that a human would have been involved later for acts of specific, targeted social engineering such as the specific phishing email sent to the CEO. While it was impossible to ascertain with precision whether or not and how the botnet/ criminal got into the network it is probable that social engineering could have played a part.

### ii. Case B: Small Company –Accounting Firm

Company Z is a small accounting firm with annual turnover of USD 2 million. In January 2019, Company Z received an invoice from Company V requesting payment of USD 20,000. Company V's normal email format was (the first name abbreviation).(the last name)@companyv.com. For example, n.nelson@companyv.com. Company Z received a fraudulent email from n.nelson@companyvv.com requesting the invoice to be paid. The fraudster had even gone so far as to register the domain name companyvv.com. While the invoice contained variant bank account details, this didn't cause any alarms on Company Z's part as many companies that they engage with have offices in different parts of the world where the bank details can change. Company V contacted Company Z in February requesting payment. At this point both Companies realised that a payment diversion had occurred. Upon further investigation it was revealed that Company V had been compromised through a mass phishing email sent to nearly all the email addresses in the company with more than one employee opening the link that downloaded malicious software onto their systems. It was further revealed that Company Z was not the only victim as a result of Company V having been compromised. Company V, however, did not alarm the criminal and instead, contacted an entity to conduct an investigation. The investigation revealed that the email had originated in the south of Nigeria. Meanwhile, payment was disguised to be going to a major bank in Norway, but was instead routed to a branch of the major Norwegian bank in Ghana which is relatively close to Nigeria, a country world-renowned for cyber fraudulent scams and incidences.[22]

### iii. Insights from Case Studies

As can be seen from the case studies, PDF affects businesses of all sizes and levels of sophistication. Spoofing was used in both cases, the first involved a text spoof and the second an email spoof. The content in the spoof texts and emails was carefully written to resemble traditional correspondence within the organisation. Most importantly, the amount targeted in both instances was very specific to the regular transactions of the organisations, and not a random amount generated. As will be seen later with ransomware, the amounts tend to be similar regardless of the size or annual turnover of the organisation.

## C. Threat Vectors

The threat vectors used to initially infiltrate an organisation's network vary greatly. Some of the more common techniques are summarised below. Additionally, threat vectors used during the actual PDF attack process are discussed.

### i. Phishing for Access

In this attack vector, a generalized email will be sent with a link to a fake site access page. This access page will be used to try and trick a user in an organization to hand over their credentials. This vector is highly generalized, and the only user specific information contained in the email would likely be the salutation, similar to the customisation in a standard mass email.

### ii. Spear-Phishing for Information

In the spear-phishing for information vector, an attacker will send an email to a specific potential target.[23] The email will look like it is from a legitimate source and will use information gained during the initial reconnaissance phase to customize the email to entice the specific user to give their credentials to an attacker in order to gain access to the target systems. This type of attack is more advanced than just using someone's name in the salutation.

[22] Muktar Bello 'Investigating Cybercriminals in Nigeria: A Comparative Study' (DPhil Thesis, University of Salford 2018) <http://usir.salford.ac.uk/id/eprint/47190/> accessed 6 May 2020.

[23] MITRE Corporation, 'Spearphishing Attachment' (*MITRE ATT&CK*, 18 April 2018) <https://attack.mitre.org/techniques/T1193/> accessed 6 May 2020.

### iii. Password Spraying

This vector is used to try commonly used passwords across many system access points in a short amount of time. The passwords tried come from lists of the most used passwords. An example would be to try and access all of the email accounts of an organization's users at the same time with a common password.

### iv. Drive-by-download

This vector allows a malware to infect users' devices by exploiting simple security flaws. Attackers place the malware often on compromised websites, then the malware automatically downloads and installs itself on the victim's device once the website is accessed.[24] Drive-by-download links are also distributed in malicious emails.

### v. System vulnerabilities

Technology giants and software vendors usually announce system vulnerabilities discovered and security patches available to mitigate security risks. Cybercriminals often exploit known vulnerabilities in unpatched system networks, providing them access to distribute ransomware payload on vulnerable devices.[25] For example, the WPA2 weakness and processor vulnerabilities.[26]

There are also threat vectors to be mindful of that are used in the actual payment diversion fraud itself. The following vectors are frequently used for PDF.

### vi. Spear-Phishing For Escalated Privileges

In this vector, spear-phishing emails can be sent from actual organization-owned email addresses in order to gain escalated privileges for the attacker. The email sent will usually have a link to a site that social engineers the user in an organization to give away their credentials.

[24] Niels Provos and others, 'All Your iFRAMEs Point to Us' (Proceedings of the 17th conference on Security symposium, July 2018).

[25] Dan Goodin, 'Serious flaw in WPA2 protocol lets attackers intercept passwords and much more' (*Ars Technica*, 16 October 2017) <https://arstechnica.com/information-technology/2017/10/severe-flaw-in-wpa2-protocol-leaves-wi-fi-traffic-open-to-eavesdropping/> accessed 22 March 2019.

[26] Mathy Vanheof and Frank Piessens, 'Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2' (Proceedings of the 24th ACM Conference on Computer and Communications Security, 2017) <https://www.krackattacks.com/> accessed 22 March 2019; 'Meltdown and Spectre' (*Meltdown Attack)* <https://meltdownattack.com/> accessed 22 March 2019.

#### vii. Email Spoofing

As previously described, email spoofing can be used by either falsifying a real email address in the 'from' line of an email through spoofing techniques, or utilising an email from an attacker-owned domain that is similar in nature to the actual organization domain being targeted, but with a single character change in the domain name.

### D. Concluding Remarks

Many of the methods used in payment diversion frauds are similar to methods and vectors exploited found in other types of cyber attacks such as ransomware which is explored below. The processes are highly automated and often involve machine-learning where human behaviour is imitated. A most recent PDF involved a deep fake where artificial intelligence was used to mimic the voice of a CEO requesting funds to be transferred to a third party.[27] We expect this type of AI enabled voice fraud to become more prevalent. To date, the authors have not seen such a toolkit available on the dark net but it is only a matter of time before one is made, and is available as an underground cybercrime tool kit and service.

## III. Ransomware

### A. What is Ransomware?

Ransomware is a type of malicious software that prevents users from accessing their system or personal files usually by locking them through encryption, and demands ransom payment in order to regain access.[28] As will be explored below, the methods and vectors have some overlap with PDF.

Ransomware, belonging to the crypto virology nest, was first introduced in 1989 and physically distributed via floppy disks at a conference event.[29] As there are no specific laws prohibiting the creation of malicious code and software, some individuals create ransomware as tools that hackers can purchase.[30] There are, however, laws that criminalise the use of prohibited

[27] Jesse Damiani, 'A Voice Deepfake Was Used to Scam a CEO Out of $243,000' (*Forbes*, 3 September 2019) <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/#63c6a1ca2241> accessed 7 May 2020.

[28] 'Ransomware' (*Malwarebytes*) <https://www.malwarebytes.com/ransomware/> accessed 7 May 2020.

[29] Ronny Richardson and Max M. North, 'Ransomware: Evolution, Mitigation and Prevention' (2017) 13(1) International Management Review 12.

[30] Thomas J. Holt, Adam M. Bossler, and Kathryn C. Seigfried-Spellar, 'Malware and Automated Computer Attacks' in *Cybercrime and Digital Forensics: An Introduction*

tools/devices. Devices have been widely defined to include algorithms which, depending on how the ransomware is written, are illegal to possess, as well as to use. Over the last three-decades, ransomware has evolved into an arsenal in the hands of cybercriminals and for as low as USD 750, cyber attackers can obtain a huge collection of ransomware to attack their victims over the internet.[31] Artificial Intelligence engineered malware is now emerging where a bot will mimic human behaviour in a way specifically designed for the target.[32]

Presently, cybercrime is listed as one of the most reported frauds, and in the majority of cases reported, cybercriminals used ransomware to obtain money from the victims.[33] Hackers primarily target sensitive information and data, deploying ransomware to a victim's device, encrypting files and information and often locking the victim out of the system.[34] There are also instances where ransomware is used as a decoy attack: while victim scrambles to pay for the decryption key for their data, the attacker often accesses the victim's data and publishes the data on illegal websites for further financial gain.[35]

Despite the fact that viruses have been around for as long as computers have, ransomware proves substantially different due to its ability to use cryptographic algorithms designed to block users' access by holding data or even the entire device hostage until a ransom is paid. This type of extortion racket with fiscal motive is unlike other malware attacks, since victims are made aware of the exploit and then given mandate directions on how to regain access. Payment often comes in bitcoins, making it easier for the perpetrators to remain unidentified.

(Routledge, 2017) 501.

[31] Kim Crawley, 'Ransomware For Sale On The Dark Web Is A Killer Bargain For Criminals' (*The Threat Report*, 12 November 2018) <https://www.thethreatreport.com/ransomware-for-sale-on-the-dark-web-is-a-killer-bargain-for-criminals/> accessed 19 March 2019.

[32] Kevin Townsend, 'IBM Describes AI-Powered Malware That can Hide Inside Benign Applications' (*Security Week*, 13 August 2018) <https://www.securityweek.com/ibm-describes-ai-powered-malware-can-hide-inside-benign-applications> accessed 7 May 2020.

[33] Nick Robinson and Tareq Hadad, 'Pulling fraud out of the shadows: A spotlight on the Middle East' (*PwC*, 2018) 30 <https://www.pwc.com/m1/en/publications/documents/economic-crime-fraud-survey-2018.pdf> accessed 7 May 2020.

[34] Jinal P. Tailor and Ashish D. Patel, 'A Comprehensive Survey: Ransomware Attacks Prevention Monitoring and Damage Control' (2017) 4 International Journal of Research and Scientific Innovation 116.

[35] Philip O'Kane, Sakir Sezer and Domnhall Carlin, 'Evolution of Ransomware' (2018) 7(5) IET Networks 321 <https://doi.org/10.1049/iet-net.2017.0207> accessed 7 May 2020.

As a self-propagating malicious program, ransomware essentially involves five stages (as illustrated in Figure 1 before the functionality of a user's device or organisation's information systems are compromised.

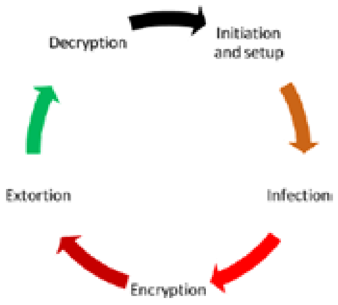

**Figure 1:** Ransomware lifecycle

### i. Initiation and setup phase

In the first stage, the cybercriminal or attacker identifies the target for the attack such as an individual, or organisation. The attacker gathers relevant information on the target from open sources (websites, social media, newspapers) to launch a successful attack. The setup may involve creating and deploying websites, emails and bogus information to lure or trap the target.

### ii. Infection phase

The second stage involves the main activities in the ransomware attack process. The attacker selects the attack medium and vector to aid the delivery of the ransomware. The internet serves as the primary medium to reach the targeted victims. Attackers often use social engineering tactics and phishing to gain access to the victim's device and network. In phishing, users usually receive spam emails marked as urgent but containing malicious links and codes. Other methods of infection include, but are not limited to, software update, drive-by-downloads, and installers. Once the target's system is infected by the malicious program, the next stage (encryption) becomes activated.

### iii. Encryption phase

In this phase, the malicious program searches the victim's device, system, or network to encrypt specific files and folders. Some ransomware encrypt system disk drives and network shared drives, and delete any backup folders and restore points. In the encryption stage, the malicious program often collects and sends details of the victim's device or system to the attacker.

### iv. Extortion phase

After the encryption process is completed, the victim usually receives an email or a prompt for the ransom payment. The victims are often given a deadline for payment to receive a decryption key to restore the data and systems back to normal, and failure to make payments will result in the total loss of data. Attackers use pseudo-anonymous methods to obtain payment from victims to prevent the authorities from tracking them. Typically this involves the use of cryptocurrencies such as Bitcoin and Monero.

### v. Decryption phase

This is the final stage in the ransomware cycle. If the targeted victim makes payment to the attacker's pseudo wallet, a decryption key is sent to the victim for data retrieval. However, decryption and restoration of the data is not guaranteed to the victim as the attackers often go back to the extortion phase to gain more from the victim.

## B. Case Study

Ransomware has grown to be one of the most advanced and destructive diabolical type of malware, able to cause worldwide catastrophes, from crippling critical infrastructure such as health, transport, and financial services to shutting down manufacturing processing plants. In 2017, the WannaCry breed of ransomware alone infected more than 2,00,000 computers in 150 countries within a day. With the advancement of ransomware and exploit kits in the hands of cybercriminals, more and more prominent attacks are witnessed on a regular basis. Herjavec[36] observed that the global annual cost of cybercrime by means of ransomware to cause damage, fraud, identity theft, and stolen personal and financial data is predicted to exceed trillions of dollars by 2021. The value of the compromised data often leaves ransomware attack victims with no choice but to pay the stated ransom to cybercriminals. The cases below provide some insights on the negative financial effect of ransomware.

The cases below involve ransomware that utilises automated software, with some use of machine-learning, but AI in the strictest sense has not

[36] Herjavec, '2019 Official Annual Cybercrime Report' (Herjavec Group, 2019) 12 <https://www.herjavecgroup.com/wp-content/uploads/2018/12/CV-HG-2019-Official-Annual-Cybercrime-Report.pdf> accessed 7 May 2020.

yet been used in ransomware, though researchers suspect that this is only a matter of time. AI models for cyberattack in the future will identify targets through facial recognition, geolocation, and voice recognition as well as mimic such human behaviours. The AI component of ransomware will most likely rest in the initial compromise of a system. Once compromised, the ransomware can sit stealthily, gathering data and waiting to declare its presence, and ask for a ransom to be paid.

### i. Case A – CryptoLocker Ransomware

CryptoLocker was first used in a cyberattack from September 2013 to May 2014. Its success led to the emergence of other ransomware variants and subsequent cyberattacks.[37] CryptoLocker was propagated via spam email with an infected attachment; primary targets were businesses and professionals. Once the victim's system was infected, encryption was executed and a ransom fee via MoneyPak or an equivalent in Bitcoin current was demanded for payment within 72 hours.[38] About 1.3% of the victims affected by the ransomware paid the ransom, but not all users received a decryption key.[39] CryptoLocker ransom payment was estimated at USD 27 million in just the first two months and had infected 2,34,000 computers by April 2014.[40]

### ii. Case B – WannaCry Ransomware

In May 2017, WannaCry ransomware was targeted at computers running Microsoft Windows operating system. The ransomware attack was wide spread infecting more than 2,00,000 systems in over 150 countries across health care, government, and telecommunication organisations.[41] The WannaCry attack lasted for a few days and was contained when a security researcher activated a kill-switch to stop the spread and locking of

[37] Josh Fruhlinger, 'Recent Ransomware Attacks Define the Malware's New Age' (*CSO*, 20 February 2020) <csoonline.com/article/3212260/recent-ransomware-attck-define-the-malwares-new-age.html> accessed 18 June 2020.

[38] Kevin Liao and others, 'Behind closed doors: Measurement and analysis of CryptoLocker ransoms in Bitcoin' (Proceedings of the 2016 APWG Symposium on Electronic Crime Research, June 2016) <https://asu.pure.elsevier.com/en/publications/behind-closed-doors-measurement-and-analysis-of-cryptolocker-rans> accessed 7 May 2020.

[39] Mark Ward, 'Cryptolocker victims to get files back for free' (*BBC*, 6 August 2014) <https://www.bbc.com/news/technology-28661463> accessed 18 June 2020.

[40] US Department of Justice, 'US Leads Multi-National Action Against "Gameover Zeus" Botnet and "Cryptolocker" Ransomware, Charges Botnet Administrator' (2 June 2014) <https://www.justice.gov/opa/pr/us-leads-multi-national-action-against-gameover-zeus-botnet-and-cryptolocker-ransomware> accessed 7 May 2020.

[41] ibid.

devices.[42] Some victims paid the ransom demanded and three months after the WannaCry ransomware attack, about £108,000 was withdrawn from the associated Bitcoin wallet.[43]

### iii. Case C – Georgia ransomware

The US has been known to be one of the most targeted countries for cyber-attacks. For example, the state of Georgia experienced ransomware cyber-attacks consecutively in 2018 and 2019. In March 2018, the city of Atlanta also experienced a ransomware attack where the attackers demanded ten Bitcoins from the government.[44] Atlanta city did not pay the ransom, but the damages and expenses to restore the systems back online resulted in millions of dollars and a long time dealing with the loss of data.[45] One year later, in 2019, Jackson County was hit with a ransomware cyberattack that crippled the IT network and systems in government offices.[46] However, unlike the city of Atlanta, Jackson County paid a hefty ransom to the attackers to obtain access to their information after the lockout.[47]

### iv. Insights from Case Studies

The case studies highly the different threat vectors and manners of escalation in ransomware. Ransomware may vary in the same ways as other forms of malware such as viruses and worms. Unlike PDF, ransomware is spread randomly from system to system and amounts tend to be similar irrespective of the size and capacity of the organisation.

[42] Sir Amyas Morse KCB, Comptroller and Auditor General, National Audit Office, United Kingdom, 'Investigation: WannaCry cyber attack and the NHS' (Department of Health, 24 October 2017) <https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf> accessed 7 May 2020.

[43] O'Kane (n 35).

[44] Lily Hay Newman, 'Atlanta Spent $2.6 M to Recover From a $52,000 Ransomware Scare' (*Wired*, 23 April 2013) <https://www.wired.com/story/atlanta-spent-26m-recover-from-ransomware-scare/> accessed 7 May 2020.

[45] Jon Fingas, 'Atlanta ransomware attack may cost another $9.5 million to fix' (*Endgadget*, 6 June 2018) <https://www.engadget.com/2018/06/06/atlanta-ransomware-attack-struck-mission-critical-services/> accessed 21 March 2019.

[46] Catalin Cimpanu, 'Georgia county pays a whopping $400,000 to get rid of a ransomware infection' (*ZD Net*, 9 March 2019) <https://www.zdnet.com/article/georgia-county-pays-a-whopping-400000-to-get-rid-of-a-ransomware-infection/> accessed 21 March 2019.

[47] Linn F. Freedman, 'Jackson County, Georgia Pays Hackers $400,000 After Ransomware Attack' (*The National Law Review*, 14 March 2019) <https://www.natlawreview.com/article/jackson-county-georgia-pays-hackers-400000-after-ransomware-attack> accessed 21 March 2019.

## C. Threat Vectors

Ransomware attack requires a vector for the covert deployment of an infection to the victim. The attack vectors for ransomware vary in complexity and effectiveness,[48] and the most prevalent ones are:

### i. Malicious emails / Social Engineering

This is one of the most common attack vectors often distributed via phishing. In some attack scenarios, an attacker employs social engineering to lure the victim into opening a malicious email attachment that will enable the execution of the ransomware payload.

### ii. Brute force - Remote Desktop Protocol

On the network level, an attacker gains admin access to server credentials with remote access. Once within the network, the attacker could exploit administrative tools and vulnerabilities to distribute and infect other devices within the network.

### iii. Exploit Kits

These are software packages used to create vulnerabilities within a system or network in order to perform malicious activities. For example, Eternal Blue was used in the 2017 WannaCry ransomware attack that infected over 2,00,000 systems globally.

### iv. Malvertising

Targeted adverts are usually displayed to potential victims based on their search history or certain web preferences. As an attack vector, malvertising displays advert with hidden malware links but mirrored as a normal advert specifically placed by a cybercriminal. Attackers often use malvertising on highly reputable websites to target their victims.[49]

[48] Aaron Zimba, 'Malware-Free Intrusion: A Novel Approach to Ransomware Infection Vectors' (2017) 15(2) International Journal of Computer Science and Information Security 317.

[49] Xinyu Xing and others, 'Understanding Malvertising Through Ad-Injecting Browser Extensions' (Proceedings of the 24th International Conference on World Wide Web, May 2015) <https://doi.org/10.1145/2736277.2741630> accessed 7 May 2020; Yuliya G. Zabyelina, 'Can criminals create opportunities for crime? *Malvertising* and illegal online medicine trade' (2016) 18(1) Global Crime 31 <https://doi.org/10.1080/17440572.2016.1197124> accessed 7 May 2020.

### v. Drive-by-download

This vector allows a malware to infect users' devices by exploiting simple security flaws. Attackers place the malware often on compromised websites, then the malware automatically downloads and installs itself on the victim's device once the website is accessed.[50] Drive-by-download links are also distributed in malicious emails.

### vi. System vulnerabilities

Technology giants and software vendors usually announce system vulnerabilities discovered and security patches available to mitigate security risks. Cybercriminals often exploit known vulnerabilities in unpatched system networks, providing them access to distribute ransomware payload on vulnerable devices.[51] For example, the WPA2 weakness and processor vulnerabilities.[52]

### vii. Network propagation

Organisations and individuals are always connected to networks to enable the seamless sharing and transfer of data. Ransomware is also capable of spreading from computer to computer over a network. On a shared network, an attack on a victim's device is easily distributed to every connected device and service within the same network. For example, the NotPetya breed of ransomware infected every machine on the Maersk global network.[53]

### viii. Propagation through shared services

Online services could also propagate ransomware. For example, infections on a home computer could easily be transferred to an office or to other connected computers if the ransomware places itself inside a shared folder.

Ransomware distribution channels are endless, and the distributors are becoming more crafty. One click could be all it takes to become a victim. Technical controls for screening and spread prevention, including having adequate backups are important to survive a ransomware attack.

[50] Provos (n 24).

[51] Goodin (n 25).

[52] Vanheof and Piessens (n 26); 'Meltdown and Spectre' (n 26).

[53] Catalin Cimpanu, 'Maersk Reinstalled 45,000 PCs and 4,000 Servers to Recover From NotPetya Attack' (*Bleeping Computer*, 25 January 2018) <https://www.bleepingcomputer.com/news/security/maersk-reinstalled-45-000-pcs-and-4-000-servers-to-recover-from-notpetya-attack/> accessed 7 May 2020; Lee Mathews, 'NotPetya Ransomware Attack Cost Shipping Giant Maersk Over $200 Million' (*Forbes*, 16 August 2017) <https://www.forbes.com/sites/leemathews/2017/08/16/notpetya-ransomware-attack-cost-shipping-giant-maersk-over-200-million/#5b32046e4f9a> accessed 7 May 2020.

## IV. Legal Framework

The Convention on Cybercrime, an agreement between member nations of the European Union, is the only international agreement in the area of cybercrime. It is unique in that it is open for signature by non-EU states. The United States, Canada, and Australia, for example, have signed and ratified the Treaty. By contrast, India has neither signed nor ratified the convention.

The convention may be divided into three key divisions: substantive law, procedural requirements, and international cooperation. All signatories to the convention must criminalize certain activities. The convention creates four main categories of substantive offences:

1. offences against the confidentiality, integrity, and availability of computer data and systems, comprising interference and misuse of devices (computer hacking offences);
2. computer-related offences, such as forgery and computer fraud;
3. content-related offences, in particular the production, dissemination, and possession of child pornography; and
4. offences related to copyright infringement.

Both socially engineered and AI enabled fraud generally involves both the computer hacking offences and computer-related fraud offences. Particularly any access, modification, or interference of a computer is criminalised. Also, misuse of devices may also be criminalised. Here devices can be defined as a hacking tool such as Zeus Malware Kit or a ransomware kit. These are algorithms and not a physical tool or kit. The 'misuse of a device' does not involve the malicious use of a hacking tool, one need only to prove intent to use the device for an illegal purpose such as fraud. This can be tricky where a device has dual purpose, but in the case of crimeware kits such as Zeus and ransomware kits, there is no dual purpose and intent is easily proven.

As seen in the examples in Sections 2 and 3, not all socially engineered frauds involve both hacking and fraud offences. They may only involve one or the other depending on the circumstances. The mere sending of a deceptive email with intent to commit fraud would not be criminalised under the Convention or under Canadian law. It would, however, be criminalised under Australian law. The Australian Criminal Code has a provision of dishonest use of a computer or device with intent to commit fraud. Of course, most jurisdictions in the world criminalise fraud, whether it is committed online or offline.

The Table below looks at the Convention provisions as well as the Canadian and Australian provisions. These two jurisdictions have been highlighted merely because of the authors' familiarity with these two jurisdictions.

| **Council of Europe Convention on Cybercrime 2001** | **Canada Criminal Framework<br>Criminal Code 1985** | **Australia Criminal Framework<br>Criminal Code 1995** |
|---|---|---|
| **Offences against the confidentiality and availability of computer data and systems** | Generally, Canada uses broad language to capture the obligations under the Convention. | Generally, Australia has very detailed provisions that address specific aspects of the Convention's obligation. What is criminalised is clear. |
| Article 2—Illegal access | Section 342.1 of the Criminal Code | Division 477 – Serious Computer Offences<br>Sections 477.1-477.3<br>Sections 478.1-4 |
| Article 3—Illegal interception | No direct equivalent | Subdivision B – Interference with telecommunications 474.3 – 474. 11<br>Division 477 – Serious Computer Offences<br>Sections 477.1-477.3<br>Sections 478.1-4 |
| Article 4—Data interference | Section 430 (1.1) of the Criminal Code | See Above. |
| Article 5—System interference | No direct Equivalent | Division 477 – Serious Computer Offences<br>Sections 477.1-477.3<br>Sections 478.1-4 |
| Article 6—Misuse of devices | Section 326 (1)(b) of the Criminal Code<br>Section 327 (1) of the Criminal Code | Section 408E (Computer hacking and Misuse) |
| **Computer-Related Fraud and Forgery** | | Part 10.8 Financial Information Offences |
| Article 7 Computer – Related Forgery | Section 366 | |
| Article 8 Computer – Related Fraud | Section 366 | Sections 480.1-480.6 |

The Convention also addresses the procedural aspects of cybercrime. The main categories here are:

1. expedited preservation of stored computer data;
2. expedited preservation and partial disclosure of traffic data;
3. production orders;
4. search and seizure of stored computer data;
5. real-time collection of traffic data; and
6. interception of content data.

In theory the procedural aspects allow collaboration between law enforcement in different jurisdictions to gather intelligence, and obtain and preserve evidence. The reality, however, is that criminals use anonymising technologies such as TOR, TAILS, and VPNs and making traceback extremely difficult. The money is equally difficult to trace as it moves from one bank to another in notable tax haven jurisdictions or moves through cryptocurrency.[54] The difficulties in traceback and cryptocurrencies are explored in the following section.

## V. Insurmountable Challenges

International law enforcement co-operate on a range of investigations and prosecutions of criminals related to cybercrime. Recent examples include the take down of two dark net markets, Hansa and AlphaBay.[55] The FBI and US Drug Enforcement Agency organised and collaborated with law enforcement from around the world to shut down AlphaBay which was in 2017 the world's largest dark net. AlphaBay boasted over 40,000 vendors and nearly a quarter of a million users/customers. Authorities arrested the mastermind and administrator of the site, Canadian Alexandre Cazes, in Thailand. Additionally, hundreds of arrests were made in countries around the world of various narcotic and weapons vendors selling on AlphaBay. In June 2017, Dutch police and Europol had secretly taken over the dark net market Hansa. At that time when AlphaBay disappeared many users and

[54] Maurushat and PhD candidate Halpin are involved with the development of a cryptocurrency database matcher and tracer technology. The technology is being developed to assist with the large growing body of investigation work with cryptocurrency fraud, as well as cryptocurrency as a money-laundering tool.

[55] Andy Greenberg, 'Global Police Spring a Trap on Thousands of Dark Web Users' (*Wired*, 20 July 2017) <https://www.wired.com/story/alphabay-hansa-takedown-dark-web-trap/> accessed 7 May 2020.

vendors flocked to competitor Hansa.[56] Later in July 2017, it was publicly announced that Dutch police had been running Hansa for a month, gathering intelligence and evidence.[57] That site was also then shut down. Hundreds of arrests of vendors were made following the takedowns.

Similarly, Interpol, along with Europol and law enforcement around the world, run operations together to take down child pornography rings, as well as anti-terrorism operations. While online fraud and computer offences fall within the jurisdiction of Interpol, there is significantly less international cooperation in this field to arrest and prosecute online fraudsters.

International organisations such as Interpol provide information about online fraud scams, monitor cases, and provide intelligence and support to national police enforcement agencies, but successful international fraud investigations of online fraud organisations is few and far between. There simply aren't the resources to conduct an international fraud investigation because those resources are spent and used elsewhere.[58]

There is often a false belief among law-makers and academics that if the right legislation is enacted, and *if enough* resources are allocated to the task, that the law can rise to the challenge and overcome a myriad of obstacles to combat cybercrime. This is, however, simply not the case for online fraud, and in particular where cybercrime is enabled by AI. The existing criminal provisions for fraud in most jurisdictions would allow for a successful prosecution of a fraudster irrespective of whether a computer was used to assist with the fraud.

After attending many conferences both within Australia and Canada representing a private cybercrime investigation firm, invariably law enforcement will ask how much money was spent on an internationally coordinated investigation. This can range between USD 2,00,000 to USD 5,00,000. Time and time again law enforcement have stated that the same investigation by law enforcement would cost ten times that amount. Below we discuss why this is the case.

[56] Andy Greenberg, 'Operation Bayonet: Inside the Sting that Hijacked an Entire Dark Web Drug Market' (*Wired*, 8 August 2017) <https://www.wired.com/story/hansa-dutch-police-sting-operation/> accessed 7 May 2020.

[57] MIX, 'Dutch Police Secretly Ran a Huge Dark Web Drug Marketplace for a Month' (*TNW*, 20 July 2017) <https://thenextweb.com/insider/2017/07/20/police-fbi-drug-dark-web-market/> accessed 7 May 2020.

[58] At the International Cybercrime Conference in Vancouver (2018) law enforcement commented that a private investigation costing USD 2,50,000 would be closer to USD 2 million if law enforcement were to undertake the same investigation.

Cybercrime investigations involve unique challenges. The challenges involve difficulty with the harmonisation of laws, jurisdictional issues, resource implications, lack of training, ambiguity in terms of how a criminal provision will be interpreted alongside human-rights protections, and, above all, a host of technical hurdles that makes tracing back to the offender difficult. Additionally, online fraud is not seen as having health and safety repercussions like other crimes, therefore, it is not prioritised. In spite of advances in machine learning, big data techniques, and artificial intelligence, attribution remains a formidable challenge.

## A. Jurisdiction

Computer crimes often involve parties located abroad. These crimes may involve people located in different jurisdictions, whether they are different states or provinces within a country or different countries altogether. Each jurisdiction may have its own laws dealing with an issue as well as its own unique set of evidence procedures in courts. Uniformity is a real problem. Successful prosecution often involves assistance and cooperation of authorities from an outside jurisdiction.[59] For a variety of reasons, some jurisdictions may or may not be willing to cooperate. Such cooperation generally must proceed through the cogs of bureaucracy in cases where time and access to good digital evidence (unaltered) is of the essence. This often means applying for warrants in multiple jurisdictions, which may translate into a loss of valuable time, and perhaps a loss of obtainable intelligence and evidence.

Private investigation firms ('**PI**'s) are less hampered with timely investigation and jurisdictional issues. If there is actionable intelligence, a PI merely picks up the phone to another PI located in that area, and contracts with them then and there to do a job immediately. This network of over 4,000 PIs world-wide operating with this type of agility makes private PIs more able to investigate online fraud. For example, the author worked on one investigation where an email tracker was sent to a spokesperson for the fraudulent company operating out of Thailand. The victim had contacted law enforcement first, but was told that they could not help her. At that point she contacted a PI who was able to act immediately on her behalf. As the victim hadn't let on that she knew that she was being defrauded, active intelligence could be gained through re-social engineering the conman. A series of email and telephone requests asking to speak to someone more senior resulted in a successful email track to a device being used in a pub in Bristol, England.

[59] For a broad in discussion of cybercrime and jurisdiction *see* Bert-Jaap Koops and Susan W. Brenner (eds), *Cybercrime and Jurisdiction: A Global Survey* (T.M.C. Asser Press, 2016).

A PI in Bristol was called to go to this pub immediately and take photos of people operating a laptop or mobile device as this is truly the only way to ascertain who the person is using a device. In this instance an identity was made by asking the pub owner a few questions. From there the case unravelled, and foreign police could be brought in for the arrests of the individuals in question. It isn't a matter of law enforcement not being involved, but a matter of when to involve them.

## B. Attribution

In cybercrime and cybersecurity, figuring out who is the person or entity responsible for an attack is the greatest singular challenge. Attribution takes three forms: who are the humans behind the incident; what devices are involved with the incident; and who may be claiming responsibility of the attack (how to verify if this is false). The greatest challenge remains in identifying and determining the physical location of the computer, and then the actual individual(s) who used the computer/network to commit a crime. As seen in the above example, a PI had to go to the pub to take a photo of the individual as they were corresponding in real time with the victim.

Let us look at another example. Police in Canada, for example, cannot obtain a warrant to wiretap someone in Mongolia, and they cannot compel an ISP in Papa New Guinea to provide data logs immediately. This type of international policing requires the cooperation of law enforcement and courts in other jurisdictions. Law enforcement could contact authorities in the location of the hacker, but cooperation may not be forthcoming. First, inter-jurisdictional investigations rely on the offence being given similar priority in both jurisdictions. For truly repugnant cases, such as child pornography, jurisdictions tend to have similar strong mandates. In the case of hacking (i.e., unauthorized access) and fraud, the priorities are often disparate.

Ironically, law enforcement have much greater capabilities and can access rich communication and fraud information that PIs cannot access. For example, law enforcement can follow two trails: the communications data trail and the financial trail. Law enforcement can access stored communication such as the content of an email or the content of a text message. Law enforcement have access to capabilities such as Cellebrite forensic tools which can bypass Apple iPhones encryption. One can only license Cellebrite if one is a law enforcement agent in a designated jurisdiction. Law enforcement can also store, access, and use metadata with great facility. The same holds true for financial information. The right tools and legislative powers exist to allow for successful prosecution, however, there are only a handful of successful international investigations of online fraud leading to arrest and prosecution. This is, again, due to lack of resources and inadequate budgets, the ability to immediately follow a lead in another jurisdiction, and the lack of law enforcement in another jurisdiction to respond to the lead with the same immediacy.

The reality is that law enforcement tend to use their resources to respond to local problems. Where there is no victim in the locale of a particular police force, priority there will not be given to an overseas investigation. Another challenge is what is known as the 'de minimus rule', whereby in order to justify valuable police resources, a certain threshold of damages must be met. The jurisdictional hurdles stem from practical considerations as well as a lack of criminalization of an act across jurisdictions.

## C. Remedies

Ironically, the main reason why using a PI is more effective than the use of law enforcement is the highly practical issue of remedy. If you lose USD 2 million it is likely that recovering the money would be your first priority. A successful arrest and prosecution resulting in prison time would be a secondary benefit. The laws in most jurisdictions, however, are designed such that a successful police operation may result in arrest, prosecution, and jail time, but no money is recovered. This is due to a number of possibilities. The first, is that some jurisdictions such as Australia have a bi-furcated approach to fraud. If a victim reports the fraud to police, and there is enough evidence to prosecute, the perpetrator in question could go to prison but the victim then has to hire a lawyer to proceed in civil proceedings in order to try to recover the lost funds. In other jurisdictions such as New York, the process of asset recovery and criminal sanction in terms of sentencing are all done at once in the court.

Often the money has been laundered in safe haven jurisdictions, or increasingly is stored as a cryptocurrency – both of which are tremendously difficult to recover funds from.[60] Or if money was miraculously recovered, the enabling legislation for proceeds of crime is inherently complex, expensive and challenging as the victim must bring a case before the court. Even when there has been a successful civil claim to recover funds, it is often the case that the defendant will claim bankruptcy. The portion of assets recovered generally is merely the tip of the iceberg. The remainder of money obtained through

[60] Cryptocurrency recovery is performed by specialist technology companies such as Cryptofound Recovery, a Silicon Valley company specialising in cryptocurrency forensics (<www.cryptogound.com> accessed 6 May 2020).

fraudulent means is nearly always located in tax havens or in untouchable cryptocurrencies.

## VI. How to Effectively Fight Online Fraud?

In jurisdictions such as Australia, fraud is handled by State law enforcement. This typically means that most successful fraud cases are ones where the criminal and victim are located in the same state. Online fraud is rarely based in one jurisdiction. The author has seen cases involving more than 32 jurisdictions. Organised online crime is sophisticated. Tackling this successfully requires both national and international coordination. In Australia at least, the Australian Federal Police (‘**AFP**’) *should* (but currently do not) take the lead on fraud cases instead of the State. For example, asking for help from overseas law enforcement must go through the designated authority under the Convention on Cybercrime. If a police officer in the State of Queensland had a lead on someone in England, a request to assist would have to go through the AFP. This process is not time-efficient whereas cybercrime leads are time sensitive.

Statistics are frightfully poor for organised fraud. Often a victim will contact law enforcement and then be told that there is nothing that they can do about it given the complexity and jurisdictional issues.[61] If an organisation lost $20,000 in a ransomware payment, this simply isn’t sufficient to warrant an investigation. But the real crime is that the details of the fraud are not captured into databases allowing fraud cases to be linked within the State, Nation, and around the world. This is very problematic. On paper a victim may only have lost USD 30,000 but collectively if the data were analysed, the same ransomware or PDF fraud may have affected hundreds of victims around the world with totals loss closer to the USD 2,00,00,00,000 mark. This is simply not captured with the way in which law enforcement collects data or chooses not to record the data accurately. Indeed there are many barriers to law enforcement sharing raw data, as well as data analytics.

Bennett-Moses and Maurushat undertook a study of data sharing amongst Australian law enforcement and intelligence agencies as part of the D2D Cooperative Research Centre. A portion of the work and findings from the study was published in an online submission to the Australian government:

> Some of the challenges are definitional. For example, different legislation will use different terms (and different definitions of the same

[61] *See* Alana M. Maurushat, ‘Botnet Badinage: Regulatory Approaches to Combating Botnets’ (DPhil thesis, University of New South Wales, 2011).

> term) to describe the object of analysis – is it data, information, communications, records, or documents? And are these physical things or digital signals or both? There are also different terms to describe the relationship between such things and particular agencies responsible for them – data might be held, in the custody of an agency, under the control of agency, in the possession of an agency, in the care of an agency, or an agency might be responsible for it or have acquired or obtained it. Again, each of these terms often comes with conflicting definitions.
>
> In addition to definitional issues, there is an issue with the assumption of much legislation about data (or equivalent term) that it is held (or equivalent term) by one entity. The question is then whether it is given to another entity and in what circumstances this is required, encouraged, permitted, or punished. However, none of this works as well with new ways of storing data – a common data platform through which multiple agencies can access data stored on one or more public or private servers does not fit easily into the existing framework.
>
> All of these issues are discussed in the report, albeit in the specific context of law enforcement information sharing. The advantages of a single Act that resolves the current confusion, dealing with all information sharing questions in a principle-based way according to a coherent set of concepts are great. Such an Act can and should recognise distinctions based on the diversity of data and circumstances, but there is no need for hundreds of separate provisions in different legislations using inconsistent concepts and definitions. For example, only a thorough review, based on existing work of the ALRC, can derive a principles-based understanding of the circumstances in which secrecy laws are appropriate. Our report included recommendations as to how the legal framework could be reworked in order to improve information sharing for law enforcement purposes. These could be combined with this project in order to improve the current complex patchwork laws rather than being excluded from scope.[62]

While the above highlights the difficulty in information sharing within Australia, there are even greater barriers to sharing information with overseas law enforcement. There may be issues of trust, having to work within the Mutual Legal Assistance Treaty framework, and budgetary restrictions.

[62] Lydia Bennett-Moses and Alana Maurushat, 'D2DCRC Information Sharing Report' (2018), cited in Lydia Bennett-Moses and others, 'Response to Issues Paper on Data Sharing and Release' (2019) UNSW Law Research Paper No.19-13 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3348816> accessed 7 May 2020.

Hiring a Private Investigation firm specialising in asset recovery is a more effective way of recording funds but this is simply not an option for most victims. As discussed prior, a typical investigation will cost between USD 2,50,000 to USD 5,00,000. A victim would have to put this money up front to run the investigation with no chance of recovery. Often multiple victims will pool their money to run the investigations but more times than not a private investigation is out of reach for the victim.

The reality is that cyber-insurance is the method that most organisations turn to when they have been the victim of PDF or Ransomware. Curiously, many cyber insurance policies do not specifically cover social engineered fraud unless a company's internal network or computer has been compromised. If, for example, an employee was contacted over the phone and was tricked to give key information to a criminal, then a deceptive email was sent which did not involve breaking in to or accessing the network or system, this is not considered 'cyber'. Or when a company becomes victim to a PDF scam due to client's compromised system, this does not always meet the definition of 'cyber' or 'computer' within some insurance policies. Careful attention must be paid to the wording of cyber insurance policies.

A new market selling decryption keys for Ransomware variants has emerged. Only a few companies offer such services, and the price to decrypt is often more than the ransom but many firms are choosing to purchase the decryption key, rather than reward those criminals behind ransomware. But decryption keys are not available for all variants of ransomware, only a select few.

One concept that is yet to be fully explored in the online fraud space is to offer a bounty for information leading to the arrest behind organised cybercrime fraud. A firm would invest their own money to investigate online fraud syndicates then receive a large portion of the funds recovered. The incentive would have to be substantial but it could prove to be an effective method down the road. How such a program would look in practice would clearly present with many significant challenges. As online fraud becomes more advanced incorporating AI enabled malware, traceback to the individuals and organisations involved in fraud will become more difficult. New methods such as bounties may be required as the technologies progress.

## VII. Concluding Remarks

This article has looked at socially engineered payment diversion fraud and ransomware from the perspective of real cases, and the experiences of the authors working in the field. The authors are working on two research projects related to Socially Engineered Payment Diversion Fraud as well as Ransomware. While the empirical findings are not yet complete from these projects, initial insights have been shared in this article.

Socially engineered payment diversion fraud and ransomware have many similarities including threat vectors and information cycles. They largely differ, however, in amount and reconnaissance. Ransomware tends to request affordable payment amounts where a company can easily see the benefit of immediate payment. These amounts range between USD 10,000 and USD 50,000. For ransomware, often a criminal or algorithm has stealthily been inside a network observing and probing for an effective way to ransom the data. PDF by contrast does not necessarily have to involve system compromise, or length periods of reconnaissance. The amounts stolen, however, have a range of between USD 5,000 to more than USD 1,00,00,000.

Law enforcement has limited capability in dealing with online organised fraud due to issues of jurisdiction, attribution, resources, and the ability to follow leads in a timely fashion. Hiring a private cybercrime investigation firm, while likely more effective in dealing with frauds involving multiple jurisdictions, is simply out of reach for many organisations. Organisations in the case of ransomware either pay the ransom or purchase the decryption key. If they have cyber insurance they will attempt to make a claim post-incident. In the instance of PDF, the amount is so substantial that a firm will want to have a full audit of its systems performed, and then implement a series of operational changes to help mitigate and prevent further instances. A firm may wish to employ a cybercrime investigation firm to assist in recovering funds. Cyber insurance might also play a role for PDF.

Moving forward into the future, the emerging field of blockchain used for logistics in supply chains is promising as is the progression towards quantum encryption and quantum decryption. Both of these methods, however, will only help prevent some forms of PDF and ransomware. Criminals are smart. They evolve to ensure a continued livelihood. Even if detection, prevention, and mitigation techniques are significantly improved, targeting the weaknesses of human beings to be socially engineered will never completely disappear.

## Annex: Essential Terms

**Adware:** Any software program in which advertising banners are displayed as a result of the software's operation. This may be in the form of a pop-up or as advertisements displayed on the side of a website, such as on Google or Facebook.

**Artificial Intelligence:** An area of computer science that emphasises the creation of intelligent machines that mimic human behaviour.

**Back door:** A method of accessing a computer program or network that circumvents security mechanisms. Sometimes a programmer will install a back door so that the programmer can accesses the program to perform security patches, troubleshoot, or monitor use. Attackers, however, can also use backdoors that they discover (or install themselves) as part of an exploit.

**Bot server and command-and-control (C&C) source:** C&C refers to the communications infrastructure of a botnet. A botnet master issues commands and exercises control over the performance of bots. Bots fetch data from a pre-programmed location, and interpret that data as triggers for action and instructions on what function to perform. The pre-programmed location is known as the bot server or C&C source. C&C is achieved by means of a bot server. The term 'server' refers to any software that provide services on request by another piece of software, which is called a client. The bot requests and the server responds. Where the client is a bot, the server is reasonably enough called a bot server. Common bot servers are IRC servers, HTTP servers, the DNS (by means of TXT records), peer-to-peer nodes, cloud nodes, and increasingly devices otherwise known as the Internet of things (e.g., Xbox).

**Bot:** A software that is capable of being invoked from a remote location in order to provide the invoker with the capacity to cause the compromised computer to perform a function. Botnets have a modular structure whereby modules (bots) may be added or taken away from each bot to add to it new exploits and capabilities. This ensures a botnet master's ability to rapidly respond to technical measures set up to infiltrate and take down the botnet.

**Botnet:** A collection of compromised computers that are remotely controlled by a bot master.

**Compromised computer:** The term 'compromised computer' is commonly used interchangeably, and in some cases wrongly, in the literature with 'zombie', 'bot', and 'bot client', which confuses hardware with software, creates inconsistency of usage, and may be confusing to users. Herein, a 'compromised computer' is a computer that is connected to the Internet (an internet is any network of any size that uses the protocol TCP/IP, and the Internet is the largest such internet) and on which a bot is installed. The computer is thus said to be compromised.

**Crypto currency:** A digital monetary currency in which encryption techniques are used to generation of units of currency which can then be verified to authorise the transfer of funds.

**Dark Net**: A subsection of the deep web – the portion of the Internet purposefully not open to public view through search engines or www protocol - where hidden networks such as Tor, VPN or TAILS are required to access the network. Dark nets are similar to underground markets where illicit goods are traded.

**Distributed Command and Control (or super botnets):** A type of botnet that draws on a small botnet comprised of fifteen to twenty bots. The botnet herders may have anywhere from 10,000 to 2,50,000 bots at their disposal but use a select few for a particular purpose. The smaller botnet is then used to issue commands to larger botnets (hence the term 'distributed command and control').

**Distributed denial of service (DDoS):** A DDoS attack is the most common form of online civil protest. A denial-of-service attack is distributed when multiple systems flood a channel's bandwidth and/or flood a host's capacity (eg, overflowing the buffers). This technique renders a website inaccessible. DDoS attacks are performed with a botnet, with several of these being used simultaneously. A DDoS attack may also be distributed by use of peer-to-peer nodes. A botnet is comprised of core elements. They are defined below for clarity and will be re-examined in more specific contexts in the analysis that follows this section.

**DNS hijacking:** DNS (domain name system) hijacking allows a person to redirect web traffic to a rogue domain name server. The rogue server runs a substitute IP address to a legitimate domain name. For example, www.alanna.com's true IP address could be 197.653.3.1, but the user would be directed to 845.843.4.1 when they look for www.alanna.com. This is another way of redirecting traffic to a political message or image.

**Dynamic DNS:** A service that enables the domain name entry for the relevant domain name to be updated very promptly, every time the IP address changes. A dynamic DNS provider enables a customer to either update the IP address via the provider's web page or using a tool that automatically detects the change in IP address and amends the DNS entry. To work effectively, the time-to-live value for the DNS entry must be set very short, to prevent cached entries scattered around the Internet serving up outdated IP addresses.

**Encryption:** It is the conversion of plain text into 'cipher text', encrypted information. Encryption acts to conceal or prevent the meaning of the data from being known by parties without decryption codes. Botnet instructions commonly use encryption. Encrypted instruction can then not be analysed, making investigation, mitigation, and prevention much more difficult. Public-key cryptography is often used. In public-key cryptography, a twin pair of keys is created: one is private, the other public. Their fundamental property is that, although one key cannot be derived from the other, a message encrypted by one key can only be decrypted by the other key.

**Exploit:** It is the implementation, in software, of a vulnerability.

**Fast flux:** A particular, dynamic DNS technique used by botnet masters whereby DNS records are frequently changed. This could be every five minutes. Essentially, large volumes of IP addresses are rapidly rotated through the DNS records for a specific domain. This is similar to dynamic DNS tactics. The main difference between dynamic DNS and fast flux is the automation and rapidity of rotation with a fast-flux botnet. Some fast-flux botnets rotate IP addresses every five minutes, and others every hour.

**Harm:** Anything that has deleterious consequences, which includes injury to persons, damage to property, financial loss, loss of value of an asset, and loss of reputation and confidence. Harm arises because a threatening event impinges on a vulnerability.

**Malware:** A simplistic definition of malware is malicious software. Malware, for the purpose of this research, is defined as potentially harmful software or a component of software that has been installed without authorization to a third-party device.

**Multihoming:** It involves the configuration of a domain to have several IP addresses. If any one IP address is blocked or ceases to be available, the others essentially back it up. Blocking or removing a single IP address, therefore, is not an effective solution to removing the content. The content merely rotates to another IP address.

**Organised crime:** A category of transnational, national, or local groupings of highly centralized enterprises run by criminals who intend to engage in illegal activity, most commonly for profit.

**Penetration/intrusion testing:** A type of information-systems security testing on behalf of the system's owners. This is known in the computer-security world as ethical hacking. There is some argument, however, as to whether penetration testing must be done with permission from a system's owners or whether a benevolent intention suffices in the absence of permission.

**Phishing:** The dishonest attempt to obtain information through electronic means by appearing to be a trustworthy entity.

**Proxy servers:** A service (a computer system or an application) that acts as an intermediary for requests from clients by forwarding requests to other servers. One use of proxy servers is to get around connection blocks such as authentication challenges and Internet filters. Another is to hide the origin of a connection. Proxy servers obfuscate a communication path such that user M connects to a website through proxy server B, which again connects through proxy server Z, whereby the packets appear to come from Z not M. Traceback to Z yields information of an additional hurdle, however, as packets also appear to come from B. Other proxy servers such as Tor are anonymous.

**Ransomware:** A type of malicious software that prevents the user from accessing or using their data (often through encrypting the data), whereby a fee must be paid or service performed before the user's data is decrypted.

**Rootkits:** Software or hardware devices designed to gain administrator-level control and sustain such control over a computer system without being detected. A rootkit is used to obscure the operation of malware or a botnet from monitoring and investigation.

**Safeguard:** A measure intended to avoid or reduce vulnerabilities. Safeguards may or may not be effective and may be subject to countermeasures.

**SQL injection:** Defacing a website involves the insertion of images or text into a website. This is often done via a SQL (structured query language) injection. A SQL injection is an attack in which computer code is inserted into strings that are later passed to a database. A SQL injection can allow someone to target a database giving them access to the website.

**TAILS:** It is a live operation system that functions from a USB stick, DVD, or external hard-drive that, once installed onto your external device, preserves your privacy and provides anonymity for online use. Essentially it forces all connections through the Tor network, then leaves little to no trace on the computer once used.

**Threat:** A circumstance that could result in harm or damage and may be natural, accidental, or intentional. A party responsible for an intentional threat is referred to as an attacker.

**Threatening event:** An instance of a generic threat (such as malicious code) that may cause harm or damage.

**Tor:** It protects you by bouncing your communications around a distributed network of relays run by volunteers all around the world. It prevents somebody from watching your Internet connection from learning what sites you visit, and it prevents the sites you visit from learning your physical location. It is described as onion routing due to the use of multiple layers of proxy servers, similar to the multiple layers of an onion. It is used by users in heavily Internet-censored countries, like China and Iran, to access blocked websites, as well as by some criminals to prevent law enforcement from traceback to the source.

**Virtual private network (VPN) service:** A network that uses a public telecommunications infrastructure (usually the Internet) to connect remote sites or users together. This connection allows secure access to an organization's network. Instead of a dedicated, real-world connection such as a leased line, a VPN uses virtual connections 'routed through the Internet from an organization's private network to the remote site or employee'. VPN is made secure through cryptographic tunnelling protocols that provide confidentiality by blocking packet sniffing and interception software.

**Virus:** A block of code that inserts copies of itself into other programs. Viruses generally require a positive act by the user to activate them. Such a positive act would include opening an email or attachment containing the virus. Viruses often delay or hinder the performance of functions on a computer, and may infect other software programs. They do not, however, propagate copies of themselves over networks. Again, a positive act is required for both infection and propagation.

**Vulnerability:** A feature or weakness that gives rise to susceptibility to a threat. Vulnerabilities exist in software and hardware.

**Worm:** A program that propagates copies of itself over networks. It does not infect other programs, nor does it require a positive act by the user to activate the worm. It replicates by exploiting vulnerabilities.

**Zero day:** An exploit or vulnerability that is exploited against a target on the day on which public awareness of the existence of the vulnerability occurs (i.e., zero days have elapsed between the awareness and the use).