1

What is Cyber Threat Intelligence and How is it Evolving?

1.1. Introduction

Today's cyberattacks have changed in form, function and sophistication during the last few years. These cyberattacks no longer originate from digital hacktivists or online thugs. Held by well-funded and well-organized threat actors, cyberattacks have transformed from hacking for kicks to advanced attacks for profit which may range from financial aims to political gains. In that aim, attacks designed for mischief have been replaced with dynamic, stealthy and persistent attacks, known as advanced malware and advanced persistent threats (APTs). The reason is due to the complexity of new technologies. As a system gets more complex, it gets less secure, making it easier for the attacker to find weaknesses in the system and harder for the defender to secure it (Schneier 2000). As a result, attackers have a first-mover advantage, by trying new attacks first, while defenders have the disadvantage of being in a constant position of responding, for example better anti-virus software to combat new malwares and better intrusion detection system to detect malicious activities. Despite spending over 20 billion dollars annually on traditional security defenses (Piper 2013), organizations find themselves faced with this new generation of cyberattacks, which easily bypass traditional defenses such as traditional and next-generation firewalls, intrusion prevention systems, anti-virus and security gateways. Those defenses rely heavily on static malware signatures or lists of pattern-matching technology, leaving them extremely vulnerable

Chapter written by Wiem TOUNSI.

to ever-evolving threats that exploit unknown and zero-day vulnerabilities. This calls for a new category of threat prevention tools adapted to the complex nature of new generation threats and attacks. This leads to what is commonly named cyber threat intelligence (CTI). CTI or threat intelligence means evidence-based knowledge representing threats that can inform decisions. It is an actionable defense to reduce the gap between advanced attacks and means of the organization's defenses. We focus specifically on technical threat intelligence (TTI), which is rapidly becoming an ever-higher business priority (Chismon and Ruks 2015), since it is immediately actionable and is easier to quantify than other TI sub-categories. TTI is also the most-shared intelligence, because of its easy standardization (Yamakawa 2014). With TTI, we can feed firewalls, gateways, security information and event management (SIEM) or other appliances of various types with indicators of compromise (IOC) (Verizon 2015), for example malicious payloads and IP addresses. We can also ingest IOC into a searchable index or just for visualization and dashboards.

Despite its prevalence, many problems exist with TTI. These are mainly related to the quality of IOC (i.e. IP address lifetime, malware signatures) and the massive repositories of threat data given by provider's databases which overwhelms their consumers (e.g. threat analysts) with data that is not always useful, which should be essential for generating intelligence. In many cases, threat feeds can simply amount to faster signatures that still fail to reach the attackers. For example, specific malicious payloads, URLs and IP addresses are so ephemeral that they may only be used once in the case of a true targeted attack.

To date, few analyses have been made on different types of TI and specifically on TTI. Moreover, very little research surveys have been reported on how new techniques and trends try to overcome TTI problems. Most of the existing literature reveals technical reports exposing periodic statistics related to the use of threat intelligence (Ponemon 2015; Shackleford 2015; Shackleford 2016), and also interesting empirical investigations for specific threat analysis techniques (Ahrend *et al*. 2016; Sillaber *et al*. 2016).

In order to develop effective defense strategies, organizations can save time and bypass confusions if they start defining what threat intelligence actually is, and how to use it and mitigate its problems given its different sub-categories.

This chapter aims to give a clear idea about threat intelligence and how literature subdivides it given its multiple sources, the gathering methods, the information life-span and who consumes the resulting intelligence. It helps to classify and make distinctions among existing threat intelligence types to better exploit them. For example, given the short lifetime of TTI indicators, it is important to determine for how much time these indicators could be useful.

We focus particularly on the TTI issues and the emerging research studies, trends and standards to mitigate these issues. Finally, we evaluate most popular open source/free threat intelligence tools.

Through our analysis, we find that (1) contrary to what is commonly thought, fast sharing of TTI is not sufficient to avoid targeted attacks; (2) trust is key for effective sharing of threat information between organizations; (3) sharing threat information improves trust and coordination for a collective response to new threats; (4) a common standardized format for sharing TI minimizes the risk of losing the quality of threat data, which provides better automated analytics solutions on large volumes of TTI.

1.2. Background

The new generation threats are no longer viruses, trojans and worms whose signatures are known to traditional defenses. Even social engineering and phishing attacks are now classified as traditional. New generation threats are multi-vectored (i.e. can use multiple means of propagation such as Web, email and applications) and multi-staged (i.e. can infiltrate networks and move laterally inside the network) (FireEye Inc. 2012). These blended, multi-stage attacks easily evade traditional security defenses, which are typically set up to inspect each attack vector as a separate path and each stage as an independent event. Thus, they do not view and analyze the attack as an orchestrated series of cyber incidents.

1.2.1. *New generation threats*

To bring new generation attacks into fruition, attackers are armed with the latest zero-day vulnerabilities and social engineering techniques. They utilize advanced tactics such as polymorphic threats and blended threats (Piper 2013), which are personalized to appear unknown to signature-based tools and yet authentic enough to bypass spam filters. A comprehensive

taxonomy of the threat landscape is done by ENISA (The European Network and Information Security Agency) in early 2017 (ENISA 2017). In the following sections, we provide some examples of these new generation threats.

1.2.1.1. *Advanced persistent threats (APTs)*

APTs are examples of multi-vectored and multi-staged threats. They are defined as sophisticated network attacks (Piper 2013; FireEye Inc. 2014) in which an attacker keeps trying until they gain access to a network and stay undetected for a long period of time. The intention of an APT is to steal data rather than to cause damage to the network. APTs target organizations in sectors with high-value information, such as government agencies and financial industries.

1.2.1.2. *Polymorphic threats*

Polymorphic threats are cyberattacks, such as viruses, worms or trojans that constantly change ("morph") (Piper 2013), making it nearly impossible to detect them using signature-based defenses. Evolution of polymorphic threats can occur in different ways (e.g. file name changes and file compression). Despite the changing appearance of the code in a polymorphic threat after each mutation, the essential function usually remains the same. For example, a malware intended to act as a key logger will continue to perform that function even though its signature has changed. The evolution of polymorphic threats has made them nearly impossible to detect using signature-based defenses. Vendors that manufacture signature-based security products are constantly creating and distributing new threat signatures (a very expensive and time-consuming proposition (Piper 2013)), while clients are constantly deploying the signatures provided by their security vendors. It is a vicious cycle which goes to the advantage of the attacker.

1.2.1.3. *Zero-day threats*

Zero-day threats are cyber threats on a publicly unknown vulnerability of an operating system or application. It is so named because the attack was launched on "day zero" or before public awareness of the vulnerability and, in many cases, before even the vendor was aware (Piper 2013). In some cases, the vendor is already aware of the vulnerability, but has not disclosed it publicly because the vulnerability has not yet been patched. Zero-day attacks are extremely effective because they can go undetected for long

periods (i.e. for months, if not years), and when they are finally identified, patching the vulnerability still takes days or even weeks.

1.2.1.4. *Composite threats*

Cyberattacks can either be classified as syntactic or semantic attacks. A combination of these two approaches is known as a composite attack or blended attack (Choo *et al*. 2007). Syntactic attacks exploit technical vulnerabilities in software and/or hardware, for example a malware installation to steal data; whereas semantic attacks exploit social vulnerabilities to gain personal information, for example scam solicitations. In recent years, progress has been made using the two approaches to realize composite attacks: using a technical tool to facilitate social engineering in order to gain privileged information, or using a social engineering means to realize a technical attack in order to cause harm to network hosts. Composite attacks include phishing attacks (also called online scams) which frequently use emails to send to carefully selected victims a plausible-looking message including a malicious attachment targeting a zero-day vulnerability. Phishing is positioned in the first three steps of the kill chain (see section 1.2.2.1). Phishing attacks forge messages from legitimate organizations, particularly banking and finance services, to deceive victims into disclosing their financial and/or personal identity information or downloading malicious files, in order to facilitate other attacks (e.g. identity theft, credit card fraud, ransomware (National High Tech Crime Unit of the Netherlands police, Europol's European Cybercrime Centre, Kaspersky Lab, Intel Security 2017)). When the attack focuses on a limited number of recipients to whom a highly personalized message is sent, the technique is named spear phishing. Phishing mostly abuses information found in social media (Fadilpasic 2016). Attackers are always on the lookout for new attack vectors for phishing including smart devices. Such devices are increasingly being used to access and store sensitive accounts and services (Choo 2011).

Obviously, the attack morphology is different depending on the aimed scenario; for example, cybercrime might use stealthy APT to steal intellectual property, while cyber war uses botnets to run distributed denial-of-service (DDoS) attacks (Skopik *et al*. 2016).

1.2.2. *Analytical frameworks*

Some analytical frameworks provide structures for thinking about attacks and adversaries to allow defenders to take decisive actions faster. For example, we name the defensive perspective of a kill chain (Hutchins *et al*. 2011) and the Diamond model used to track attack groups over time. Other standardized frameworks are developed in section 8.4.

1.2.2.1. *Steps of the kill chain defensive perspective*

Kill chain, first developed by Lockheed Martin in 2011 (Hutchins *et al*. 2011), is the best known of the CTI frameworks. It is a sequence of stages required for an attacker to successfully infiltrate a network and exfiltrate data from it (Barraco 2014). By breaking up an attack in this manner, defenders can check which stage it is in and deploy appropriate countermeasures.

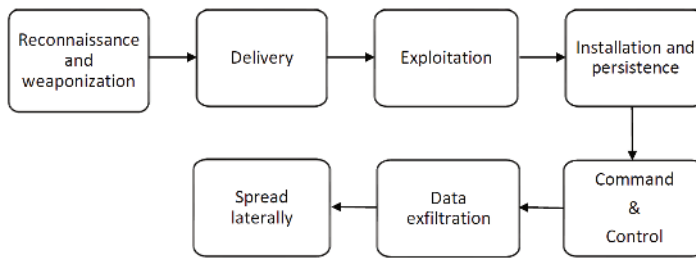

Figure 1.1. *Typical steps of multi-vector and multi-stage attacks by Lookheed Martin's kill chain*

– *Reconnaissance and weaponization*: the reconnaissance consists of research, identification and selection of targets, often by browsing websites (e.g. conference proceedings, mailing lists and social relationships), pulling down PDFs or learning the internal structure of the target organization. The weaponization is realized by developing a plan of attack based on opportunities for exploitation.

– *Delivery*: this consists of the transmission of the weapon to the targeted environment. It is often a blended attack delivered across the Web or email threat vectors, with the email containing malicious URLs (i.e. phishing attack). Whether it is an email with a malicious attachment or a hyperlink to a compromised website or an HTTP request containing SQL injection code, this is the critical phase where the payload is delivered to its target.

– *Exploitation*: most often, exploitation targets an application or operating system vulnerability, but it can exploit the users themselves or leverage an operating system feature that auto-executes a code.

– *Installation and persistence*: a single exploit translates into multiple infections on the same system. More malware executable payloads such as key loggers (i.e. unauthorized malware that records keystrokes), password crackers and Trojan backdoors could then be downloaded and installed. Attackers have built in this stage long-term control mechanisms to maintain persistence into the system.

– *Command and control (C&C)*: as soon as the malware is installed, a control point from organizational defenses is established. Once its permissions are elevated, the malware establishes communication with one of its C&C servers for further instructions. The malware can also replicate and disguise itself to avoid scans (i.e. polymorphic threats), turn off anti-virus scanners, or can lie dormant for days or weeks, using slow-and-low strategy to evade detection. By using callbacks from the trusted network, malware communications are allowed through a firewall and could penetrate all the different layers of the network.

– *Data exfiltration*: data acquired from infected servers are exfiltrated via encrypted files over a commonly allowed protocol, for example FTP or HTTP, to an external compromised server controlled by the attacker. Violations of data integrity or availability are potential objectives as well.

– *Spread laterally*: the attacker works to move beyond the single system and establishes long-term control in the targeted network. The advanced malware looks for mapped drives on infected systems, and can then spread laterally into network file shares.

Typically, if you are able to manage and stop an attack at the exploitation stage using this framework, you can be confident that nothing has been installed on the targeted systems, and triggering a full incident response activity may not be needed.

The kill chain is a good way for defending systems from attacks, but it has some limitations. One of the big criticisms of this model is that it does not take into account the way many modern attacks work. For example, many phishing attacks skip the exploitation phase and instead rely on the victim to open a document with an embedded macro or by double-clicking on an attached script (Pace *et al*. 2018). But even with these limitations, the

Kill Chain is a good baseline to discuss attacks, and find at which stage they can be stopped and analyzed.

1.2.2.2. *The Diamond model of intrusion analysis*

The Diamond model was created in 2013 at the Center for Cyber Intelligence Analysis and Threat Research (CCIATR). It is used to track adversary groups over time rather than the progress of individual attacks. The simplest form of the Diamond model is shown in Figure 1.2.

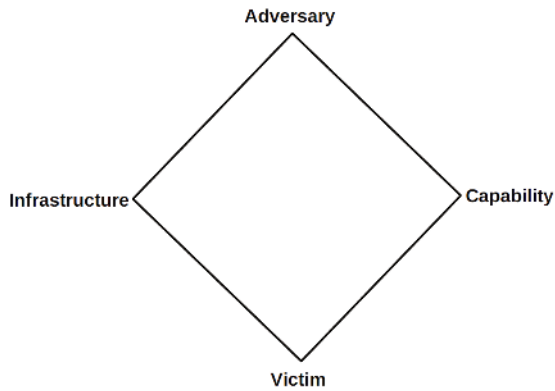

Figure 1.2. *The Diamond model of intrusion analysis*

The Diamond model classifies the different elements of an attack. The diamond for an adversary or a group is not static, but evolves as the adversary changes infrastructure and targets and modifies its TTPs (tactics techniques and procedures). The Diamond model helps defenders to track an adversary, the victims, the adversary's capabilities and the infrastructure the adversary uses. Each point on the diamond is a pivot point that defenders can use during an investigation to connect one aspect of an attack with the others (Pace *et al*. 2018).

One big advantage of the Diamond model is its flexibility and extensibility. It is possible to add different aspects of an attack under the appropriate point on the diamond to create complex profiles of different attack groups. These aspects of an attack include: phase, result, direction, methodology and resources.

This model requires time and resources. Some aspects of the model, especially infrastructure, change rapidly. If the diamond of an adversary is not constantly updated, there is a risk of working with outdated information.

1.3. Cyber threat intelligence

Cyber threat intelligence, also known as threat intelligence, is any evidence-based knowledge about threats that can inform decisions (McMillan 2013), with the aim of preventing an attack or shortening the window between compromise and detection. CTI can also be information that, instead of aiding specific decisions, helps to illuminate the risk landscape (Chismon and Ruks 2015). Other definitions exist, for example, in Steele (2014) and Dalziel (2014). A more rigorous one (Dalziel 2014) states that CTI is an information that should be relevant (i.e. potentially related to the organization and/or objectives), actionable (i.e. specific enough to prompt some response, action or decision) and valuable (i.e. the information has to contribute to any useful business outcome). CTI supports different activities, namely security operations, incident response, vulnerability and risk management, risk analysis and fraud prevention (for more details, see Pace *et al*. (2018)). Depending on the intended activities, the sources of CTI may differ.

1.3.1. *Cyber threat intelligence sources*

Cyber threat intelligence can be generated from information collected from a variety of sources (Holland *et al.* 2013). These commonly include internal sources (i.e. firewall and router logs or other local sensor traffic, such as honeynets, which are groups of interactive computer systems mostly connected to the Internet that are configured to trap attackers), or external sources, such as government-sponsored sources (i.e. law enforcement, national security organizations), industry sources (i.e. business partners), Open Source Intelligence (OSINT i.e. public threat feeds such as Dshield (2017), ZeuS Tracker (2017), social media and dark web forums) and commercial sources (i.e. threat feeds, Software-as-a-Service (SaaS) threat alerting, security intelligence providers).

External sources could provide structured or unstructured information, whereas internal sources are known to provide structured information as it is generated by technical tools. Structured sources are technical, meaning all information from vulnerability databases or threat data feeds, which are machine parsable and digestible and so their processing is simple. Unstructured sources are all that is produced by natural language, such as what we find in social media, discussions in underground forums, communications with a peer, or dark webs. They require natural language processing and machine learning techniques to produce intelligence. Table 1.1 presents these sources with required technologies to process information and transform it into intelligence.

	Internal sources	**External sources**	
	Structured (mainly)	***Structured***	***Unstructured***
Example	Firewall and router logs, honeynets	Vulnerabilities databases, IP blacklists and whitelists, threat data feeds	Forums, news sites, social media, dark web
Technologies for collecting and processing	Feed parser	Feed/web scraper, parser	Collection: crawlers, feed/web parsers Processing: Natural Language Processing (NLP), machine learning

Table 1.1. *Threat intelligence sources*

After collecting and processing threat information, several initiatives encourage threat information sharing, such as incident response teams and international cooperation (CERTs, FIRST, TF-CSIRT) (Skopik *et al*. 2016), and information sharing and analysis centers (ISACs) (ENISA 2015).

1.3.2. *Cyber threat intelligence sub-domains*

With different sources of threat intelligence and the activities that make use of it, it is useful to have subdivisions to better manage the gathered information and to focus efforts. TI can be categorized into sub-domains. Ahrend *et al*. (2016) divide TI into formal and informal practices to uncover and utilize tacit knowledge between collaborators. It depends on the collaborators' form of interaction. Gundert (2014) and Hugh (2016) categorize TI as strategic and operational depending on the form of analysis used to produce it.

In Chismon and Ruks (2015) and Korstanje (2016), a more refined model divides threat intelligence into four distinct domains: strategic threat intelligence, operational threat intelligence, tactical threat intelligence and technical threat intelligence. This subdivision is also known as the four levels of intelligence analysis (Steele 2007b). It was originally used in a military context as the model of expeditionary factors analysis that distinguishes these four levels (Steele 2007a). In what follows, our study follows the last subdivision. Table 1.2 summarizes the four domains.

– *Strategic threat intelligence* is high-level information consumed by decision-makers. The purpose is to help strategists understand current risks and identify further risks of which they are yet unaware. It could cover financial impact of cyber activity or attack trends, historical data or predictions regarding the threat activity. As a result, a board needs to consider and target possible attacks, in order to weigh risks and allocate effort and budget to mitigate these possible attacks. Strategic TI is generally in the form of reports, briefings or conversations.

– *Operational threat intelligence* is information about specific impending attacks against the organization. It is initially consumed by high-level security staff, for example security managers or heads of incident response team (Chismon and Ruks 2015). It helps them anticipate when and where attacks will take place.

– *Tactical threat intelligence* is often referred to as tactics, techniques and procedures. It is information about how threat actors are conducting attacks (Chismon and Ruks 2015). Tactical TI is consumed by incident responders to ensure that their defenses and investigation are prepared for

current tactics. For example, understanding the attacker tooling and methodology is tactical intelligence that could prompt defenders to change policies. Tactical TI is often gained by reading technical press or white papers, communicating with peers in other organizations to know what they are seeing attackers do, or purchasing from a provider of such intelligence.

– *Technical threat intelligence (TTI)* is information that is normally consumed through technical resources (Chismon and Ruks 2015). Technical TI typically feeds the investigative or monitoring functions of an organization, for example firewalls and mail filtering devices, by blocking attempted connections to suspect servers. TTI also serves for analytic tools, or just for visualization and dashboards. For example, after including an IOC in an organization's defensive infrastructure such as firewalls and mail filtering devices, historical attacks can be detected by searching logs of previously observed connections or binaries (Chismon and Ruks 2015).

	Strategic	**Operational**	**Tactical**	**Technical**
Level	High	High	Low	Low
Audience	The board	Defenders	Senior security management; architects	Security Operation Center staff; incident response team
Content	High level information on changing risks	Details of specific incoming attacks	Attackers' tactics, techniques and procedures	Indicators of compromise
Time frame	Long term	Short term	Long term	Immediate

Table 1.2. *Threat intelligence sub-domains*

From their definitions, strategic and tactical threat intelligence are gainful for a long-term use, whereas operational and technical threat intelligence are profitable for a short-time/immediate use. In case technical IOC are for

short time use, a key question is: how long we can expect those indicators to remain useful? In the next section, we deal with TTI in more detail.

1.3.3. *Technical threat intelligence (TTI)*

Defenders should not only be aware of threat actors and the nature of attacks they are facing, but also be aware of the data fundamentals associated with these cyberattacks, known as indicators of compromise (IOC). IOC are closely linked to TTI, but are often confused with intelligence. IOC are an aspect that enables the production of intelligence. The feeds by themselves are just data. By conducting the analysis with the internal data intelligence which is relevant to the organization, an actionable decision is able to recover from any incident (Dalziel 2014). IOC are commonly partitioned into three distinct categories (Ray 2015): network, host-based indicators and email indicators.

– *Network indicators* are found in URLs and domain names used for command and control (C&C) and link-based malware delivery. They could be IP addresses used in detecting attacks from known compromised servers, botnets and systems conducting DDoS attacks. However, this type of IOC has a short lifetime as threat actors move from one compromised server to another, and with the development of cloud-based hosting services, it is no longer just compromised servers that are used, but also legitimate IP addresses belonging to large corporations.

– *Host-based indicators* can be found through analysis of an infected computer. They can be malware names and decoy documents or file hashes of the malware being investigated. The most commonly offered malware indicators are MD5 or SHA-1 hashes of binaries (Chismon and Ruks 2015). Dynamic link libraries (DLLs) are also often targeted, as attackers replace Windows system files to ensure that their payload executes each time Windows starts. Registry keys could be added by a malicious code, and specific keys are modified in computer registry settings to allow for persistence. This is a common technique that malware authors use when creating Trojans (Ray 2015).

– *Email indicators* are typically created when attackers use free email services to send socially engineered emails to targeted organizations and individuals. Source email address and email subject are created from

addresses that appear to belong to recognizable individuals or highlight current events to create intriguing email subject lines, often with attachments and links. X-originating and X-forwarding IP addresses are email headers identifying the originating IP address of (1) a client connecting to a mail server, and (2) a client connecting to a web server through an HTTP proxy or load balancer, respectively. Monitoring these IP addresses when available provides additional insight into attackers.

Spam is the main means to transport malicious URLs and malware. These are later wrapped in the form of spam and phishing messages (i.e. phishing is positioned in the first three steps of the kill chain.) Phishing attacks forge messages from legitimate organizations to deceive victims into disclosing their financial and/or personal identity information or downloading malicious files, in order to facilitate other attacks). Spam is mainly distributed by large spam botnets (i.e. devices that are taken over and form a large network of zombies adhering to C&C servers (ENISA 2017)). Obfuscation methods (Symantec 2016) were observed in 2015 and continued in 2016 to evade the detection of this type of attack. These methods could be the expedition of massive amounts of spam to a wide IP range to reduce the efficiency of spam filters or the usage of alphanumeric symbols and UTF-8 characters to encode malicious URLs.

1.4. Related work

Cyber threats and attacks are currently one of the most widely discussed phenomena in the IT industry and the general media (e.g. news) (iSightPartners 2014). Figure 1.3(a) shows Google results for cyber "threat intelligence", particularly in terms of research publications, and Figure 1.3(b) shows Google results for "indicators of compromise" in the threat landscape generally and in terms of research publications particularly, in the last five years. These numbers are taken year on year. Even though an exponential interest in threat intelligence and IOC fields is seen, we observe a gap between the evolution of cyber threat intelligence activities and related research work.

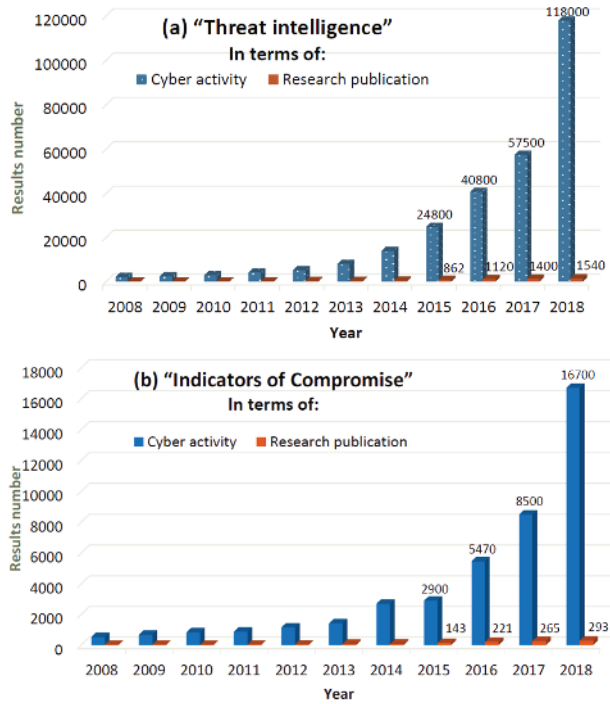

Figure 1.3. *Trend of "threat intelligence" and "indicators of compromise" in cyber activity from the last ten years. For a color version of this figure, see www.iste.co.uk/tounsi/cyber.zip*

Actually, a large number of threat intelligence vendors and advisory papers are found describing very different products and activities under the banner of threat intelligence. The same conclusion is observed with TTI category via the indicators of compromise. However, few research studies have been done to examine and identify characteristics of TI and its related issues. It is also noteworthy that in recent years, significant research progress has been made in this field.

Regarding surveys related to our work, most of them show yearly new trends and statistics which are relevant to strategic intelligence (Ponemon 2015; Shackleford 2015; Shackleford 2016). On the research side, a significant body of work has been dedicated to threat intelligence sharing issues (Moriarty 2011; Barnum 2014; Burger *et al*. 2014; Ring 2014; Skopik *et al*. 2016). Many guidelines, best practices and summaries on existing

sharing standards and techniques have been published (e.g. Johnson *et al*. 2016). In contrast, less research has been devoted to areas like TTI problems and how to mitigate them.

This work complements the aforementioned research work by separating TI categories. It specifically analyzes TTI problems per type (i.e. problems of information quantity over quality and specific limitations related to each type of IOC). Then, it shows how to mitigate them. We also survey the reasons behind not sharing threat information with peers and present solutions to share this information by avoiding either attack or business risks for organizations. We show how a common standardized representation of TTI improves the quality of threat information which improves automated analytics solutions on large volumes of TTI suffering from non-uniformity and redundancy. Finally, we evaluate TTI tools which aim to share threat intelligence between organizations.

In the following section, we start by describing the main reasons for not sharing TI.

1.5. Technical threat intelligence sharing problems

The benefits of collective sharing and learning from extended and shared threat information are undeniable. Yet, various barriers limit the possibilities to cooperate. In this section, we detail some of these benefits and expose the reasons for not sharing threat information.

1.5.1. *Benefits of CTI sharing for collective learning*

Many organizations and participants today agree on the importance of threat information sharing for many reasons. First, the exchange of critical threat data has been shown to prevent potential cyberattacks and mitigate ongoing attacks and future hazards. According to Bipartisan Policy Center (2012), leading cybercrime analysts recognize that public-private cyber information sharing can speed identification and detection of threats. Thus, if organizations are able to find an intruder in his/her active phases, they have a greater chance of stopping the attacker before data is stolen (Zurkus 2015).

In addition, threat sharing is a cost-effective tool in combating cybercrime if properly developed (PERETTI 2014; Ponemon 2014). In Gilligan *et al*. (2014), a study on the economics of cyber security identified a number of "investment principles" for organizations to use in developing data security programs with high economic benefit. One of these principles is the participation in multiple cyber security information-sharing exchanges. Advantages of sharing also include a better situational awareness of the threat landscape, a deeper understanding of threat actors and their TTPs, and a greater agility to defend against evolving threats (Zheng and Lewis 2015). This has been proved in a recent survey (Ponemon 2015), where 692 IT and IT security practitioners were surveyed across various industries. Results reveal that there is more recognition that the threat intelligence exchange can improve an organization security posture and situational awareness. More broadly, sharing threats improves coordination for a collective learning and response to new threats and reduces the likelihood of cascading effects across an entire system, industry, sector or sectors (Zheng and Lewis 2015). Many attacks do not target a single organization in isolation, but target a number of organizations, often in the same sector (Chismon and Ruks 2015). For example, a company can be damaged when a competing business's computers are attacked, since the information stolen can often be used against other organizations in the same sector.

1.5.2. *Reasons for not sharing*

Despite the obvious benefits of sharing threat intelligence, a reluctant position in reporting breaches is observed. The issue was seriously highlighted at a pan-European level when ENISA, the EU's main cyber security agency, published a report (ENISA 2013) in 2013, intentionally capitalizing the word "SHARE". The report warned around 200 major CERTs in Europe that "the ever-increasing complexity of cyberattacks requires more effective information sharing" and that organizations were not really involved in doing so. In its last report on threat landscape published in early 2017 (ENISA 2017), ENISA continues to recommend sharing information as a mitigation vector for malwares. Authors recommend the development of methods for the identification and sharing of *modus operandi* without disclosing competitive information.

Many concerns are deterrent to participation in such a sharing initiative. In Table 1.3, we identify by order of importance ten major reasons for not sharing threat information.

Fearing negative publicity is one of the main reasons for not sharing threat information which could result in a competitive disadvantage (Richards 2009; Choo 2011; Peretti 2014; Chismon and Ruks 2015); for example, competitors might use the information against victimized organization. In some sectors, even a rumor of compromise can influence purchasing decisions or market valuations (Bipartisan Policy Center 2012).

Legal rules and privacy issues are also cited among the most important reasons for not sharing (ENISA: European Union Agency for Network and Information Security 2013; PERETTI 2014; Murdoch and Leaver 2015; Skopik *et al*. 2016). Organizations may be reluctant to report an incident because they are often unsure about what sort of information can be exchanged to avoid legal questions regarding data and privacy protection. In the same country legal rules may not be the same for the collaborating parties. Affiliation to a specific sector, for example, might force adherence to specific regulations (ENISA 2006). Regarding international cooperation, confidence between cooperating teams while handling sensitive information is most of the time prevented by international regulations that limit the exchange and usage of such information. Teams working in different countries have to comply with different legal environments. This issue influences the ways the teams provide their services and the way they treat particular kinds of attacks, and therefore limits the possibilities to cooperate, if not making cooperation impossible (Skopik *et al*. 2016).

Quality issues are one of the most common barriers to effective information exchange, according to different surveys realized on CERTs and other similar organizations (ENISA 2013; Ring 2014; Ponemon 2015; Sillaber *et al*. 2016). Data quality includes relevance, timeliness, accuracy, comparability, coherence and clarity. For example, many interviewees report that a great deal of what is shared is usually a bit old and thus not actionable. It is also not specific enough to aid the decision-making process.

Untrusted participants is also cited in recent surveys (ENISA 2013; Murdoch and Leaver 2015; Ponemon 2015) among the crucial obstacles to effective communication between organizations.

1	Fearing negative publicity	Richards (2009), Choo (2011), Peretti (2014), Chismon and Ruks (2015)
2	Legal rules, privacy issues	ENISA (2013), Peretti (2014), Murdoch and Leaver (2015), Skopik *et al*. (2016)
3	Quality issues	ENISA (2013), Ring (2014), Ponemon (2015), Sillaber *et al*. (2016)
4	Untrusted participants	ENISA (2013), Murdoch and Leaver (2015), Ponemon (2015)
5	Believing that the incident is not worth sharing	Choo (2011), Ring (2014), Chismon and Ruks (2015)
6	Budgeting issues	Ring (2014), Skopik *et al*. (2016)
7	Natural instinct to not share	Ring (2014)
8	Changing nature of cyberattacks	Ring (2014)
9	Unawareness of the victimized organization about a cyber incident	Choo (2011)
10	Believing that there is a little chance of successful prosecution	Choo (2011)

Table 1.3. *Reasons for not sharing*

Some interviewees in ENISA (2013) have pointed that trust is undermined when only a few parties are active in a sharing program, without getting much in return from the other parties. Murdoch and Leaver (2015) explain that the conflict involved in the need for sharers to keep anonymity while ensuring that recipients still trust the information (i.e. even when the

source is unknown) is a barrier to participation in a threat intelligence sharing platform.

Believing that the incident is not worth sharing is also commonly stated (Choo 2011; Ring 2014; Chismon and Ruks 2015). The victimized organizations simply deal with the incident internally believing that it is not serious enough to warrant reporting it to an external party including law enforcement and other competent agencies.

Budgeting issues are reasons to limit building a valuable level of cooperation (Ring 2014; Skopik *et al*. 2016). Some interviewees have stated that qualified real-time threat intelligence is typically expensive to receive/share. They think that they are mainly dedicated to big organizations. Yet third party providers generally offer a discount to their TI platform subscribers for their willingness to share threat information in addition to its consumption (Piper 2013).

The natural instinct of organizations to not share is another problem of sharing (Ring 2014). In some organizations, there is still the perception of a blame culture (i.e. if something happens, then obviously it is somebody's fault and he/she needs to pay the price). Thus, people are naturally reticent about advertising an incident too widely.

The changing nature of cyberattacks, which are becoming increasingly personalized, is highlighted (Ring 2014). Even though organizations succeed in sharing data on an attack, especially when these organizations fall in the targeting scope of a given adversary (see the Waking Shark II exercise (Keeling 2013)), the issue of personalized attacks do not help them to defend themselves. This sets a different kind of intelligence requirement.

The ignorance of the victimized organization about a cyber incident is another reason for not sharing threat information (Choo 2011). When asked, analysts indicate that they had not experienced any such incidents. Yet the organization has been attacked one or more times.

Believing that there is a little chance of successful prosecution (Choo 2011) discourages organizations from reporting an incident to law enforcement and other competent agencies.

It is worth noting that some of these aforementioned concerns have been alleviated by recent government measures (PERETTI 2014) (e.g. legal rules). However, in all cases, organizations should understand any potential risks associated with exchanging TI and take steps to mitigate such risks. The good news is that organizations are taking strides in sharing threat data. Now the question is: how useful are these shared threats and for how long will they remain worthy of alerting/blocking?

1.6. Technical threat intelligence limitations

While a decade ago no one other than government agencies were talking about threat intelligence, as shown in Figure 1.3, organizations looking to have technical threat intelligence are now overwhelmed with a massive amount of threat data (Chismon and Ruks 2015; Zurkus 2015), leaving them with the huge challenge of identifying what is actually relevant. Thus, a problem of quantity over quality has been developed.

1.6.1. *Quantity over quality*

Most security teams cannot use their threat data in a valuable way because there is just too much. The daily dump of indicators seen as suspicious on the Internet provides information on approximately 250 million indicators per day (Trost 2014), which allows consumers to pivot around and glean their own intelligence. The brain power needed to analyze at the speed at which the threat data is produced is, thus, not humanly possible (Zurkus 2015).

Timeliness of information is also very important when protecting against aggressive attackers and zero-day exploits. According to Ponemon (2013), 57% of surveyed IT professionals state that the intelligence currently available to their enterprises is often out of date, making it difficult for them to understand the motivations, strategies and tactics of attackers and to locate them. In a more recent survey of IT professionals (Ponemon 2014), the authors report that most of the threat information the IT professionals received was not timely or specific enough (i.e. actionable) to meet their perceived need. Finally, according to a recent survey (Ponemon 2015), threat intelligence needs to be timely and easy to prioritize. In this latter survey, 66% of respondents who are only somewhat or not satisfied with current approaches explain that the information is not timely. On the other

hand, 46% complain that the threat information is not categorized according to threat types. This has been proved in Ring (2014), where the author explains that qualified real-time threat intelligence is typically expensive; otherwise, most available commercial and open-source threat intelligence products are not effective enough, or provide information that is out of date.

1.6.2. *IOC-specific limitations*

Let us now deal in more detail with the limitations of each type of indicator of compromise as categorized in section 1.3.3. These TTI types are: network indicators, host-based/malware indicators and email indicators.

1.6.2.1. *Network indicators*

A number of network indicators can be collected as TTI (as shown in section 1.3.3). In some cases, attackers use different nodes to conduct attacks on a targeted victim. In other cases, they use the same node for multiple victims. Thus, an IP address that has been observed by others, functioning as a C&C (Command and Control) node, can be a useful indicator (see section 1.2.2.1 for more details). However, attackers often use different IP addresses, changing C&C nodes as they are discovered or as victimized computers become unavailable. Regarding domain names, a malware will attempt to connect to a domain name, which can then be pointed to the IP address the attacker is currently using. It is also considered a useful indicator when a malware uses a hard-coded domain. However, it is quite common for malware to use a domain generation algorithm to avoid the need to connect to the same domain twice. In such a case, a domain name has little value as an IOC (Chismon and Ruks 2015). A stark detail in Verizon (2015) illustrates the value of such indicators. The most important experiments conducted in this field are to determine (1) the IP addresses' cumulative uniqueness (i.e. the overlap between the IP address feeds), (2) the time needed for an attack to spread from one victim to another and (3) the time validity of the IP address.

1.6.2.1.1. Cumulative uniqueness observations

For six months, Niddel (Pinto and Sieira 2015) combined daily updates from 54 different sources of IP addresses and domain names tagged as malicious by their feed aggregator TIQ-Test (Niddel Corp. 2014). The

company then performed a cumulative aggregation (i.e. if ever two different feeds mentioned the same indicator throughout the six-month experimental period, they would be considered to be in overlap on this specific indicator). To add some context to the indicator feeds being gathered, Niddel separated them into groups of inbound feeds (i.e. information on sources of scanning activity and spam/phishing email) and outbound feeds (i.e. information on destinations that serve either exploit kits or malware binaries, or even locations of C&C servers). The results show significant overlap only in the inbound feeds, which is in some way expected as every feed is probed and scanned all the time. However, despite every feed being generally subjected to the same threats, the overlap in outbound feeds is surprisingly small, even with a long exposure period of six months. This result suggests that either attackers are using huge numbers of IP addresses (i.e. organizations would need access to all threat intelligence indicators in order for the information to be helpful, which is a very hard task) or only a minority of IP addresses contained within the feeds are of intelligence value. It is likely that the truth is a mixture of both explanations (see experience in time validity observations).

1.6.2.1.2. Time spread observations

Experiments on attacks observed by RiskAnalytics (Verizon 2015) display some pretty interesting and challenging results: 75% of attacks spread from Victim 0 to Victim 1 within 24 hours and over 40% hit the second organization in less than an hour. These findings put quite a bit of pressure on the security community to collect, vet and distribute IOC very quickly in order to maximize the collective preparedness. Let us assume that indicators are shared quickly enough to help subsequent potential victims. The next question that needs to be answered is how long we can expect those indicators to remain valid (i.e. malicious, active and worthy of alerting/blocking).

1.6.2.1.3. Time validity observations

RiskAnalytics has already studied the question of IP address validity. Figure 1.4 shows how long most IP addresses were on the block/alert list. The graphic is restricted to seven days of outbound IP address feeds.

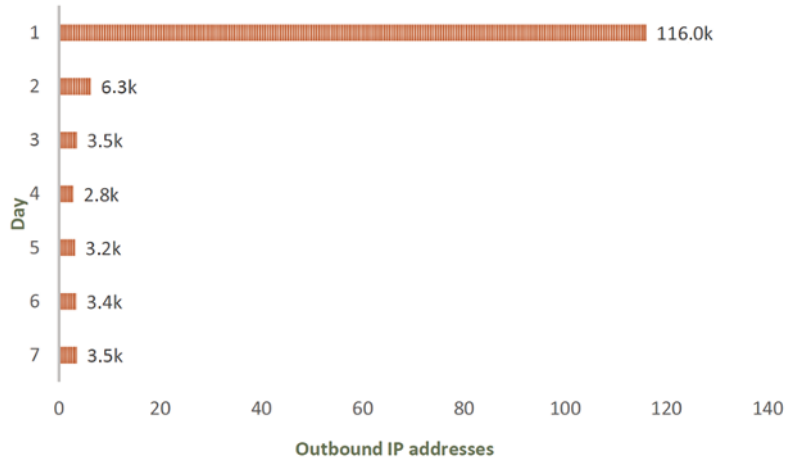

Figure 1.4. *Count of indicators by days as observed in Verizon (2015) in at least one feed*

While some IP addresses remain valid for some time, most do not last even a day. These findings track well with Niddel's cumulative uniqueness observations where the overlap in outbound feeds is very small. According to Verizon (2015), these results reflect an urgent need to share quickly: "the faster you share, the more you theoretically will stop". Let us recall that these are results from one data source which is geared towards threats of a more opportunistic, high-volume and volatile nature (e.g. brute forcing and web application exploits), rather than more slow targeted attacks. With targeted attacks, sharing IOC faster is not always useful (see section 1.8).

1.6.2.2. *Malware indicators*

A growing sophistication in the evolution of malware has been noted (Choo 2011). Knowing that modified malware does not require great skill or resources, attackers reuse malware to keep ahead of the anti-malware industry and security professionals. They adapt their "products" over time, employing many different obfuscation techniques. The simplest way is to change a single bit in the binary, and then a different hash will be obtained. Attackers can use open-source tools and make more complex modifications to change the hashes. The Zeus bot malware creator kit is an example of an easy-to-use toolkit. It can be bought or found for free on some underground forums (Falliere and Chien 2009) with detailed instructions on how to use such kits. Any individual, including one with limited programming or

hacking skills, could use such kits to create sophisticated malware or customize them to his own needs and launch advanced attacks. A study realized by Symantec (Fossi *et al*. 2010) has shown nearly 90,000 unique variants of the basic Zeus toolkit in 2009, which was the second most common new malicious code family observed in the Asia Pacific/Japan region in that year. The widespread availability of such a toolkit facilitates cybercrime. Consequently, there is a marked increase in the number of amateur cyberattackers who make their pocket money from distributing spam or selling stolen credentials and information (Choo 2011). Indicators such as created registry keys or file artifacts can be more useful for threat intelligence as they are less commonly changed by attackers, even though it is still possible to give dropped files a random or pseudorandom component in their names.

1.6.2.3. *Email indicators*

A large number of attacks start with a phishing or spear phishing attack, containing either a document exploit or simply a malware disguised as something benign. Thus, email indicators can provide useful threat intelligence. Attackers often ensure that emails are either targeted or generic (Choo 2011). Sharing generalist feeds of spam email subjects will be less useful than details of phishing emails sent to similar organizations.

1.7. Cyber threat intelligent libraries or platforms

A concept that has emerged recently is the use of threat intelligence libraries, also called threat intelligence platforms (Poputa-Clean and Stingley 2015). Threat intelligence platforms produce data and information, which human analysts can use to produce actionable threat intelligence (RecordedFuture 2017). These libraries solve collection and storing problems of TTI and facilitate sharing with other organizations in the threat intelligence space. They also allow an organization to detect attacks within logs and packet captures. These libraries particularly store indicators and seek links between them (Chismon and Ruks 2015). They are generally large repositories that often use big data technologies (e.g. graph analysis and data warehousing) to draw links between types of TTI, allowing quicker response to detected threats, as well as a historical record of an IOC. The threat library could be seen as the reference system for IOC (Poputa-Clean and Stingley 2015) since it allows the defender to track for a basic observable (e.g. a checksum of a file or a signature), as well as enriching it (e.g. by

metadata on the file, notes, campaigns using the file itself and its source). Thus, organization security is no longer dependent on having seen the threat before. It can evaluate any traffic based on the collective knowledge (Williamson 2016) of all the threats that came before it.

1.7.1. *Benefits of CTI libraries based in the cloud*

The cloud is seen as the proper place to host the CTI library, since cloud computing provides computation, storage and distributed capability in support of big data processing used by this library. The TI library is a part of a cloud threat intelligence system (Lightcyber 2016), which is constantly updated with recent IOC. The expected result behind this is, once one entity has detected a new threat in its initial attack state, all other entities connected to the system are protected within minutes. Many call this collective immunity (Piper 2013). But still, organizations should define what cloud model they want to use to be safe and to build trust upon. Whether they are private or public, threat libraries automate in the long term some of the analyst activities (Trost 2014). Their most-known benefits could be seen after a couple of months or a year, when organizations can automatize the inverse engineering by using machine learning techniques. For example, the libraries make it possible to have information associated with one group and to give its evolution, or to know whether attackers are learning from their mistakes, or to know if attackers are changing their TTP and to build an attacker profile. In addition, these libraries allow an operational usage, which helps to define the best metrics by determining what is the best source of intelligence for the organization (i.e. private sources, commercial sources, OSINT) and who is giving the best pool of information to dedicate time and resources. Today, several open-source projects and commercial enterprises are gaining popularity as TTI platforms as they offer a more organized storage of IOC and an improved context around the alerts (Poputa-Clean and Stingley 2015). The most notable open-source intelligence libraries are evaluated in section 1.9.

1.7.2. *Reluctance to use cloud services*

It is hard to deny the benefits of cloud services to host threat intelligence data and networks. However, many businesses are still uncomfortable with the idea of a cloud-based infrastructure. Even though some public cloud

providers dedicate infrastructure and services to their customers, offering a virtual private cloud, these businesses want to store their TI-related data in a privately controlled datacenter, where they control at any time where and with whom their data reside.

Some organizations such as healthcare and financial services industries are dealing with data that require more advanced security than what cloud providers can offer. Others have a problem of visibility. In this case, integrating an internal private cloud solution facilitates connection and interaction with other organizations. This could be seen as just an extension of the corporate datacenter which allows internal and external threat data (i.e. OSINT, other partners) to provide context and relevance while the most sensitive data remain safe behind the corporate firewall. Finally, trust issues often remain the most difficult to overcome for IT executives when considering whether or not to move TI into a cloud-based solution maintained by a third-party provider. In fact, trust is difficult to quantify. It depends on the specific needs of the organization, the cloud provider's overall reputation and localization (e.g. European or not, for regulations reasons), and the kinds of service-level agreements to have (OrbIT-Peaple 2016).

1.8. Discussion

1.8.1. *Sharing faster is not sufficient*

As seen before, a variety of security vendors and open-source providers now offer a wide assortment of threat feeds of the latest indicators of compromise (Williamson 2016). The idea behind these threat feeds is generally the same. As attackers are getting faster, security providers find a way to quickly aggregate and share the latest threats that have been noticed. Timeliness of threat information is very important when protecting against aggressive attackers and zero-day exploits, but in many cases, threat feeds of TTI can simply amount to faster signatures that still fail to track the attackers. In fact, a key failing of TTI is that it is relatively simple for an attacker to target a specific organization in a way that ensures no pre-existing indicators are available. For example, specific network indicators may only be used once in the case of a true targeted attack. In addition, a large percentage of malware used in breaches were reported to be unique to the organization that was infected (Shackleford 2015;

Verizon 2015). Clearly, if a threat is only used once, as for targeted attacks, faster sharing of IOC alone is not going to solve the problem. Actually, targeted attacks need a targeted defense as well (Chismon and Ruks 2015).

To defend against this new trend of personalized attacks, organizations need to focus not only on gathering and sharing threat data across their industry sector, but also on their individual threat analysis and incident response (Ring 2014). Obviously, they cannot protect themselves if they do not know what they are protecting against and who their adversaries are (Pepper 2015; Zurkus 2015). To realize such need, internal audit should be done regularly to understand organization's internal and external vulnerabilities. The objective is to make assumptions about what the attacker can do and to get an initial response, a one step forward by focusing on investigating specific devices and attacks.

1.8.2. *Reducing the quantity of threat feeds*

The other important concern is about the large amounts of data sold as TTI which lack contextual information. Certainly, anything that leads to the discovery of an incident is worthwhile, but in most cases, context is key. Additional context includes indicator role, kill chain stage, originating MD5, malware family and/or adversary group (Trost 2014). Adding, for example, the context in which previously detected attacks have taken place enables a wider audience to make a broader defensive capability. The heart of the issue is that the vast majority of information included in threat feeds is made to answer a question to a particular test. If the question on the test is changed, then the data cease to be useful (Williamson 2016). In such case, an atomic indicator has its own life which has no value in being shared with others.

Facing this challenge, the following solutions could be worth some consideration. Security teams need to contextualize the threat data they collect with the specific internal vulnerabilities and weaknesses (Bellis 2015). For this purpose, they should select the data they collect, or build/purchase large analytics platforms to cope with the quantity of data. There are massive dedicated teams, cleaning the Internet, looking at the targeted attacks, analyzing staff and trying to find associations. This process can also be automated using techniques of artificial intelligence. For example, machine learning approaches are used to build a base knowledge which will be able to infer new relationships (i.e. associations) between

entities at hand or predict new events. Following the same idea, it is suggested in Ring (2014) to use managed security services, as an increasing number of organizations start to outsource this area.

Regarding the huge flow of malware variants which is gaining access to networks and computer systems or reaching organizations' honeypots, it is impossible to handle them individually. Identifying the mutations of malware variants is essential in order to recognize those belonging to the same family. Data science and machine-learning models are looking to deliver entirely new ways of searching malware. Instead of taking a 1-for-1 approach where each threat is mapped to a signature and/or IOC, data science models are analyzing a huge number of threats, to learn what they all have in common. Methods of malware analysis, detection, classification and clustering should be either automated or designed in such a way that makes future automation possible (Ghanaei *et al*. 2016). As for new research work, in Ghanaei *et al*. (2016), the authors propose a supervised learning method based on statistics to classify new malware variants into their related malware families. In the same vein, VirusBattle (Miles *et al*. 2014) is a prototype system that employs state-of-the-art malware analyses to automatically discover interrelationships among instances of malware. This solution analyzes malware interrelations over many types of malware artifacts, including binaries, codes, dynamic behaviors and malware metadata. The result is a malware interrelation graph. Cuckoo (Guarnieri *et al*. 2016) and Malheur (Rieck 2013) are well-known open-source platforms that automate the task of analyzing malicious files in a sandbox environment. Cuckoo uses both a static analysis (i.e. code or binary analysis) and a dynamic analysis (i.e. behavior analysis) (Oktavianto and Muhardianto 2013) whereas Malheur uses a dynamic analysis. To identify a malware family using Cuckoo, one can create some customized signatures that can be run against the analysis results in order to identify a predefined pattern that might represent a particular malicious behavior. This requires a local specialist to investigate the results and classify the analyzed samples. On the other hand, Malheur uses machine learning to collect behavioral analysis data inside sandbox reports and categorizes malware into similar groups called "clusters" (Rieck *et al*. 2011). Malheur builds on the concept of dynamic analysis, which means that malware binaries are collected in the wild and executed. The execution of each malware binary results in a report of recorded behavior. Malheur analyzes these reports for discovery and discrimination of malware classes using machine learning.

Well-known public sandboxes such as Cuckoo and Malheur have now become highly detectable by malware (Issa 2012; Ferrand 2015). Once such systems are made publicly available, malware authors try to protect themselves and to evade detection by checking whether they are in an emulated environment or in a real one (e.g. by checking the execution time). To face this issue, new research is currently being carried out. For example, in Ferrand (2015), the author shows that with a few modifications and tricks on Cuckoo and the virtual machine, it is possible to prevent malwares from detecting whether they are being analyzed, or at least make this detection harder.

1.8.3. *Trust to share threat data and to save reputation concerns*

Effective sharing requires trust. Since shared threat data might be sensitive (e.g. reveal that an organization has been attacked), organizations will be reluctant to share when they are not in a trusted environment. This is proved in some studies conducted on the economic cost of sharing security data, which have demonstrated that sharing can result in a loss of market due to negative publicity (Campbell *et al*. 2003; Cavusoglu *et al*. 2004). To avoid such consequences, techniques for fine-grained and context-based access control are critical to protect confidentiality and privacy of data (Tounsi *et al*. 2013). Shared threat data could also be contaminated by malicious activity and contain erroneous information. In such a case, establishing trust at least requires authentication of transmissions and encryption of content. To avoid the consequences of identity revelation, anonymous sharing is another solution that provides participants a channel in which they can communicate anonymously. In a recent work (Dunning and Kresman 2013), the authors have developed an algorithm to anonymously share private data between participants.

Trust is also important on another level. It is generally unwise to allow threat actors to learn what you know about them, lest they change their methods. Thus, closed and trusted groups can enable deeper sharing than would otherwise be possible.

Generally, the more a group can trust its members and the security of information within the group, the more effective the sharing tends to be (Chismon and Ruks 2015). However, a certain level of trust in the group should be guaranteed. If a participant believes there is more consumption of

the threat information than sharing in the network, the motivation to share information will rapidly decline. To address this issue, some research work has been initiated. In Cascella (2008), the game theory approach using the prisoner's dilemma is employed to model the interactions of rational and selfish nodes in distributed systems. The study shows that by incepting a reputation system, it is possible to distinguish good players (threat sharers) and bad players (threat consumers). In Seredynski *et al*. (2007), a sanction mechanism that makes a decision to discard/forward packets is proposed based on an evolving genetic algorithm. The aim is to enhance trust between several nodes transmitting data packets. However, in voluntary threat sharing mechanisms, the use of sanction and punishment would not be very interesting. Other research has shown that instead of punishing, encouraging good behavior increases the likelihood of participants being more involved in a sharing program in the long run. For example, participants that receive social approval can have a significant positive impact on cooperative behavior (Cook *et al*. 2013). It is also shown that having organizations involved in the quality assurance process improves the cooperation among participants and increases the level of trust (Gerspacher and Lemieux 2010). Finally, Furnell *et al*. (2007) conclude that competitive advantage of threat intelligence can be gained for whichever side employs social factors better, involving human, social and organizational matters that are mostly uncharted on the computer security research agenda. For example, assuming that face-to-face interactions usually occur in a trusted environment (MITRE Corporation 2012), the one-to-one human contacts can be one of the simplest, yet most effective and trusted sources of actionable information (Chismon and Ruks 2015).

1.8.4. *Standards for CTI representation and sharing*

Sharing security knowledge between experts across organizations is an essential countermeasure to recent sophisticated attacks, and organizations can benefit from other organizations' experiences to build a collective knowledge (Williamson 2016).

Organizations have traditionally shared threat information using *ad hoc* solutions such as phone calls, encrypted emails or ticketing systems. More recently, they used portals and blogs, a trend of building interconnected communities with associated platforms to exchange threat information semi-automatically (Sillaber *et al*. 2016). Latent semantic analysis (LSA)

(Landauer *et al*. 1998) was used to find semantically related topics in a web blog corpus. Important keywords of each topic are assigned quantitative measure through probabilistic LSA (PLSA) (Hofmann 1999). The results prove the efficiency of this approach to broadly search security-related news in massive web blogs. However, this approach is limited because of the limitation of web blogs in representing threat scenarios in a real-time and structured manner. Li *et al*. (2007) focus on the problem of true threat identification where network security data are managed at distributed locations. The authors provide several means of finding correlations between alerts arriving from distributed components. The major limitation of this work is once more the lack of standardization as alert data need to be converted to a uniform representation given the multiple TTI sources of information. A concept that has emerged is the use of threat intelligence libraries, also called threat intelligence platforms (Poputa-Clean and Stingley 2015). These libraries were designed to solve the collection and storing problems of TTI and to facilitate sharing threat information with other organizations. However, efficient automation and collection from a diverse set of products and systems require structured and standardized threat information representation (Barnum 2014; Wagner *et al*. 2016), which is expected to be expressive, flexible, extensible, machine-parsable and human-readable (Barnum 2014; Heckman *et al*. 2015).

Several efforts have been made to facilitate threat information sharing in a standardized manner. IODEF (Danyliw *et al*. 2007), RID (Moriarty 2012), STIX (Structured Threat Information eXpression), TAXII (Trusted Automated eXchange of Indicator Information) (Barnum 2014), OpenIOC (Open Incident of Compromise) (Mandiant 2017), CybOX (Cyber Observable Expression) (MITRE 2017e), VERIS (Vocabulary for Event Recording and Incident Sharing) (Verizon 2017), CAPEC (Common Attack Pattern Enumeration and Classification) (MITRE 2017c), MAEC (Malware Attribution and Enumeration Characterization) (MITRE 2017b) and ATT&CK (Adversarial Tactics, Techniques & Common Knowledge) (MITRE 2017a) are popular examples of such standardized efforts. Despite these initiatives of standardization, an interesting survey from SANS (Shackleford 2015) indicates that in 2015, only 38% of organizations were using TI data in standard formats and well-known open-source toolkits. In order to select the right standard for a particular use case, Burger *et al*. (2014) provide an agonistic framework in which standards can be evaluated and assessed. In the following, we briefly examine the aforementioned standards.

STIX and TAXII have appeared as improvement initiatives to IODEF and RID, where RID and TAXII are the transport protocols for IODEF and STIX respectively. Formerly developed by the MITRE corporation, STIX and TAXII are sponsored by the Department of Homeland Security (DHS) office of cyber security and communications. They have been introduced to combine human and machine data for sharing information.

Today, STIX is a commonly used standard even though it is very complex to implement (Kampanakis 2014). STIX is modular and can incorporate other standards efficiently (Burger *et al*. 2014). The STIX architecture is composed of eight core cyber-threat concepts: campaigns, indicators, observables, TTP (tactics, techniques and procedures), incidents, ThreatActors, ExploitTargets and courses of action.

STIX can embed CybOX, IODEF and some OpenIOC extensions, in addition to XML namespaces, extensions for YARA rules (Google Inc. *et al*. 2017), Snort rules (The Snort Team 2017) and non-XML bindings (i.e. using JSON). STIX uses CybOX as a representation of Observables. CybOX is a schema for the specification, capture and characterization of observable operational events. STIX can also include IODEF in place of the IncidentType extension and OpenIOC extensions in place of CybOX, to express non-standard Indicator patterns.

XML namespaces that STIX can embed are the MITRE CAPEC, MAEC and ATT&CK, to cite a few. CAPEC schema attributes characterize how cyber threats are executed and provide ways to defend against these threats. MAEC schema provides a standardized language about malware based on artifacts, behaviors and attack patterns. ATT&CK was released as a framework of adversary post-compromise techniques (Strom 2016). It describes patterns to characterize adversarial behavior on the network and endpoints to achieve its objectives in a standard way. Since CAPEC enumerates a range of attack patterns across the entire cyberattack life cycle (i.e. not just techniques used during post-compromise), the CAPEC ID references are added to the attack pattern descriptions in ATT&CK. For malware researchers using YARA for regular expression matching and analysis or for communities whose interest is intrusion detection using Snort, there are extensions for YARA and Snort rules supported by STIX (Kampanakis 2014). YARA is an engine and language for scanning files and memory blocks. When a rule matches a pattern, YARA presumes to classify the subject according to the rule's behavior. Snort is an open-source packet

analyzer with intelligence to match rule sets and trigger actions when there are expression matches (Burger *et al*. 2014). Finally, VERIS system is a characterization of cyber incidents after they have occurred. It is intended for strategic trend analysis, risk management and metrics.

These multiple efforts to obtain different data ontologies for threat information sharing are not without disadvantages. These ontologies often overlap and do not offer a solution to the entire community (Burger *et al*. 2014). There could be duplications and gaps in the threat information ontologies in different communities, which lead to a duplication of effort for effective collaboration. Thus, there is a need for participants to have a common language and toolset to facilitate sharing and threat analytics.

1.8.5. *Cloud-based CTI libraries for collective knowledge and immunity*

1.8.5.1. *Using private/community cloud solutions*

There is no right or wrong answer when choosing to maintain CTI in the cloud via a third-party cloud provider or in a private collaborative solution. Organizations may find that some threat intelligence data have low impact and are relatively easy to transit within the cloud whereas other more critical data may be best kept on site. It all depends on the business, the data that the organization possesses and the comfort level of having a third party managing the risk.

The choice could also be for financial reasons. Private cloud often requires significant capital investment that covers all systems management, patching and future upgrades of hardware and software which are supported by the provider in public cloud solution. Facing such issue, the community cloud-based TI could be the best solution, as each member organization in a community cloud may host some portion, or applications, that all organizations in a community can leverage. However, it is worth noting that the deployment of cloud community requires an extensive and deep long-term relationship between multiple organizations in order to build, govern and operate from a community cloud (Bond 2015). Finally, many technical leaders (Ellison 2014) advocate a hybrid model to threat intelligence-driven security where both coexist (Rouse 2015). In such a case, multiple private/community and/or public cloud providers could interconnect.

1.8.5.2. *Being aware of major security concerns in the cloud*

Some security organizations have released their research findings on major security concerns related to the cloud to assist companies interested in joining cloud computing and security data storage. The aim is to encourage them to make a wise decision while being fully aware of the associated risks. This also encourages new customers to ask the appropriate questions and consider getting a security assessment from a neutral third party before committing to a cloud provider (Khorshed *et al*. 2012).

Since June 2008, the security farm Gartner published a report (Heiser and Nicolett 2008), where it identifies seven specific security issues that customers should raise with providers before selecting the service. The specific issues are privileged user access, regulatory compliance, data location, data segregation, recovery, investigative support and long-term viability.

In November 2009, the European Network and Information Security Agency (ENISA) published another research document for risk and recommendation in cloud computing (Catteddu and Hogben 2009), which is until now referenced by many organizations and certifications (Cloud Security Alliance 2017a). The document lists eight important cloud-specific risks which are: loss of governance, lock-in, isolation failure, compliance risks, management interface compromise, data protection, insecure or incomplete data deletion and malicious insiders. They have also discussed risk management and provided recommendations.

Cloud Security Alliance (CSA) has published via the Top Threats Working Group their most recent research findings on the top threats of cloud computing in February 2016 (Cloud Security Alliance 2016). The purpose is to provide organizations with an up-to-date, expert-informed understanding of cloud security concerns in order to identify major risks and make educated risk-management decisions regarding cloud adoption strategies. In this recent edition of the report, experts identified the following 12 critical issues in cloud security, ranked in order of severity per survey results: data breaches, weak identity, credential and access management, insecure APIs, system and application vulnerabilities, account hijacking, malicious insiders, advanced persistent threats (APTs), data loss, insufficient due diligence, abuse and nefarious use of cloud services, denial of service, and shared technology issues.

Finally, CSA released version 4.0 of its security guidance in cloud computing in December 2017, in which fourteen areas of concerns are identified and categorized into six modules (Cloud Security Alliance 2017b). These are as follows: cloud computing concepts and architectures, governance and enterprise risk management, legal issues, contracts and electronic discovery, compliance and audit management, information governance, management plane and business continuity, infrastructure security, virtualization and containers, incident response, application security, data security and encryption, identity, entitlement, access management, security as a service and related technologies. The documents have quickly become the industry-standard catalogue of best practices for secure cloud computing.

The purpose of all these research works is to assist cloud providers as well as their potential customers in identifying the major risks and to help them decide whether or not to join in cloud infrastructure, and to proactively protect them from these risks.

1.9. Evaluation of technical threat intelligence tools

Now that organizations have found ways to collect huge amounts of information from a wide variety of sources using threat libraries, they are in need of tools to manage the flow of information and convert it into knowledge and actions. Although the existing TI tools need more sophistication (Shackleford 2016), they have been able to achieve a level of maturity that enables organizations to start filtering and sharing information effectively (Brown *et al*. 2015). There are several open-source projects and commercial enterprises offering products to access threat intelligence. These solutions mainly aim at content aggregation and collaborative research such as IBM X-Force Exchange (IBM 2017), Alien-vault OTX Pulse (AlienVault 2007), Recorded Future (Pace *et al*. 2018) and Crowdstrike intelligence exchange (CrowdStrike Inc. 2014). Other solutions are focused on providing TI advanced management options with the possibility of having private instances. These include EclecticIQ (ElecticIQ 2017), Threat-stream (Anomali 2017), ThreatQuotient (ThreatQuotient 2017), ThreatConnect (Threatconnect 2017), MISP (Andre *et al*. 2011), CRITS (MITRE 2017d), Soltra Edge (Soltra 2017), CIF v3 (also called bearded-avenger) (CSIRT Gadgets 2016), IntelMQ (European CERTs/CSIRTs 2014) and Hippocampe (CERT Banque de France 2017). We are focusing on recent open-source

tools that offer advanced management options, specifically IntelMQ, Hippocampe and Threatelligence (Phillips 2014). This evaluation complements the previous comparative discussion in Tounsi and Rais (2018), where other free and/or open-source tools can be found.

1.9.1. *Presentation of selected tools*

Some of the aforementioned tools have begun to gain popularity as they promise a more organized storage of IOC and a powerful parsing of the latter. We name IntelMQ, Hippocampe and Threatelligence as examples.

These tools are presented and evaluated according to different functional dimensions (see section 1.9.2). Technical information related to these tools is taken from various sources: official tools sites, white papers, research articles and live interactions with some of the tools' authors.

IntelMQ (European CERTs/CSIRTs 2014) is a community-driven initiative called IHAP (Incident Handling Automation Project) which was conceptually designed by European CERTs units during several InfoSec events in 2014. The solution was designed for CERTs to improve the incident handling processes, using a message queue protocol, to collect and process security feeds, pastebins and tweets.

Hippocampe (CERT Banque de France 2017) is an open-source project released in 2016 by the CERT of the Banque de France. It aggregates feeds from the Internet in an ElasticSearch cluster. It allows organizations to query it easily through a REST API or a Web UI. It is based on a Python script which fetches URLs corresponding to feeds, and parses and indexes them. It is part of the Hive project which is designed to make life easier for SOCs, CSIRTs, CERTs and any information security practitioner dealing with security incidents.

Threatelligence (Phillips 2014) is a cyber-threat intelligence feed collector, using ElasticSearch, Kibana and Python. It was created in 2014 by a professional and very implicated developer in cyber security. It fetches TI data from various custom or public sources into ElasticSearch while enriching them slightly (i.e. looking for geographic locations). It automatically updates feeds and tries to further enhance data for dashboards.

The dashboards that are built using Kibana are used to display data and facilitate searching through them.

1.9.2. *Comparative discussion*

The market maturation is encouraging all the CTI vendors and open-source initiatives to adopt a common set of features. The important features include (Poputa-Clean and Stingley 2015): (1) integration with other security tools (e.g. via an API), (2) data enrichment (i.e. the possibility of integrating other intelligence platforms, for example MISP (Malware Information Sharing Platform) (Andre *et al*. 2011)), (3) integration with other tools created by the same provider and (4) sharing TI features.

Taking into account the aforementioned criteria, we focus our evaluation on the presence of Web API, indexation with MISP (as a standard OSINT feed), interaction with other tools from the same provider, IOC storage and updating, and Web User Interface (UI) capabilities. Table 1.4 summarizes this evaluation.

Criteria/Tool	Hippocampe	IntelMQ	Threatelligence
Project status	Young. Active since 2016	Mature. Active since 2014	Young. Inactive since 2014
Main language	Python	Python	Python
Built in Web API	Yes. Available to retrieve IOCs	No	No
MISP indexing	Not supported, but in the roadmap	A “Collector Robot” for MISP is available	No
Interaction with other products from the same provider	Works seamlessly with Cortex (TI analyzer) and TheHive (incident response tool)	No	No
IOC storage	Uses ElasticSearch: – IOCs are saved in ElasticSearch – IOCs are organized by feed sources	No database required: – Using Redis[1] as data store – IOCs are stored in a JSON file by default. Each IOC from a source corresponds to a JSON record in the file[2]	Uses ElasticSearch: – IOCs are saved in ElasticSearch – IOCs are organized by attack type

IOC updating	For an old IOC record in the database, update only the "last appearance" field. If a new IOC is listed by two feed sources, the IOC will be added to ElasticSearch for each feed	Not clear. The bot "Deduplicator" could filter all duplicated IOC	The same old IOC in the database will be overridden. There is no "First appearance" and "Last appearance" fields
Web UI capabilities	Display the number of IOC by type or by feed Display each feed collector working status (last running time, number of all IOC obtained, number of new IOCs)	Display collector, parser and expert relationship graph. Allow addition or deletion of bots with UI. Allow monitoring and control of bots (start, stop, reload, restart, configure)	N/A

1 https://redis.io/.

2 IntelMQ provides other storage choices, such as ElasticSearch and PostgreSQL.

Table 1.4. *Evaluation of threat intelligence tools*

1.10. Conclusion and future work

As organizations continue to move to the cloud, indicators of compromise (IOC) are also changing. Thus, the impact of cyber security to an organization goes beyond the classic bounds of the technical computer domain. Access to data, their linkage with email accounts, web applications and documents stored in a variety of platforms and mobile devices stretch the need of an organization to protect their data. We have seen that cyber threat intelligence (CTI) has many advantages in supporting several security activities. Discovering covert cyberattacks and new malware, issuing early warnings and selectively distributing TI data, are just some of these advantages. We have given a clear definition of threat intelligence and how literature subdivides it. We have focused on TTI, which is the most-consumed intelligence, and the major problems related to it. We have found that on one hand, timeliness of information is very important when protecting against zero-day exploits. On the other hand, fast sharing of IOC is not sufficient when dealing with targeted attacks, since specific network and host-based indicators may only be used once by the attacker in a true targeted attack. In this case, targeted defenses are also needed where security teams within organizations have to collect and filter threat data with a focus

on their internal vulnerabilities and weaknesses. We surveyed new research works, trends and standards to mitigate TTI problems and delivered the most widely used sharing strategies based on trust and anonymity so that participating organizations can do away with the risks of business leak. Clearly, the more a group can trust its organization members, the more effective the sharing tends to be. However, before building community-based organizations, it is worth considering some common factors that are related to business process, stability and cooperation policy. Finally, as security data are shared between participants, aggregated from different sources and linked to other data already present in the datasets, the number of errors will increase. Consequently, a standardized format of threat and a common vocabulary for extra data entry minimizes the risks of data quality issues and provides better automated analytics solutions on large volumes of threat data.

While scalability and accuracy remain an ultimate need for producing actionable IOC, approaches to the collection, processing and analysis of all kinds of intelligence have traditionally relied heavily on the capability of humans to understand references, filter out noise and ultimately make a decision about actions that need to be taken. Today's proliferation of sophisticated cyber attacks and consequently the massive number of indicators of compromise make it paramount to automate at least both information collection and processing, considering the diverse sources of threat data. Artificial intelligence is the best candidate to realize this objective, especially after the advent of cloud computing made access to supercomputer capabilities affordable and opened doors to artificial intelligence being applied to more and more complex problems.

1.11. References

Abuse.ch. (2017). Zeus Tracker [Online]. Available: https://zeustracker.abuse.ch [Accessed January 2017].

Ahrend, J.M., Jirotka, M., and Jones, K. (2016). On the collaborative practices of cyber threat intelligence analysts to develop and utilize tacit Threat and Defence Knowledge. *International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, IEEE, 1–10.

AlienVault (2007). AlienVault open threat exchange. [Online]. Available: https://www. alienvault.com/open-threat-exchange [Accessed January 2017].

Andre, D., Dereszowski, A., Dulaunoy, A., Iklody, A., Vandeplas, C., and Vinot, R. (2011). MISP: Malware Information Sharing Platform [Online]. Available: http://www. misp-project.org/ [Accessed January 2017].

Anomali (2017). [Online]. Available: https://www.anomali.com/product/threatstream [Accessed January 2017].

Barnum, S. (2014) Standardizing cyber threat intelligence information with the structured threat information expression (STIX). *MITRE Corporation*, 11, 1–22.

Barraco, L. (2014). Defend like an attacker: Applying the cyber kill chain, [Online]. Available: www.alienvault.com/blogs/security-essentials/defend-like-an-attacker-applying-the-cyber-kill-chain.

Bellis, E. (2015). The problem with your threat intelligence. White paper, Kenna Security, July.

Bipartisan Policy Center (2012). Cyber security task force: Public-private information sharing, National Security Program, July.

Bond, J. (2015). Planning and Architecture. In *The Enterprise Cloud* [Online]. O'Reilly Media, Inc. Available: https://www.safaribooksonline.com/library/view/the-enterprise-cloud/9781491907832/ch01.html.

Brown, S., Gommers, J., and Serrano, O. (2015). From cyber security information sharing to threat management. *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, ACM, pp. 43–49.

Burger, E.W., Goodman, M.D., Kampanakis, P., and Zhu, K.A. (2014). Taxonomy model for cyber threat intelligence information exchange technologies. *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security*, ACM, pp. 51–60.

Campbell, K., Gordon, L.A., Loeb, M.P., and Zhou, L. (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security*, 11(3), 431–448.

Cascella, R.G. (2008). The "value" of reputation in peer-to-peer networks. *5th IEEE Consumer Communications and Networking Conference, CCNC 2008*, IEEE, 516–520.

Catteddu, D. and Hogben, G. (2009). Benefits, risks and recommendations for information security. *European Network and Information Security*, November.

Cavusoglu, H., Cavusoglu, H., and Raghunathan, S. (2004). How should we disclose software vulnerabilities. *Proceedings of Workshop on Information Technology and Systems*, pp. 243–248.

CERT Banque de France (2017). Hippocampe [Online]. Available: https://github.com/TheHive-Project/Hippocampe [Accessed April 2018].

Chismon, D. and Ruks, M. (2015). Threat intelligence: Collecting, analysing, evaluating. MWR Infosecurity, UK Cert, United Kingdom.

Choo, K.-K.R. (2011). The cyber threat landscape: Challenges and future research directions. *Computers & Security*, 30(8), 719–731.

Choo, K.-K.R., Smith, R.G., and McCusker, R. (2007). *Future Directions in Technology-enabled Crime: 2007–09*. Australian Institute of Criminology, Canberra, Australia.

Cloud Security Alliance (2016). The treacherous 12 – Cloud computing top threats in 2016 [Online]. Available: https://downloads.cloudsecurityalliance.org/assets/research/top-threats/Treacherous-12_Cloud-Computing_Top-Threats.pdf.

Cloud Security Alliance (2017a). Certificate of cloud security knowledge [Online]. Available: https://cloudsecurityalliance.org/education/ccsk.

Cloud Security Alliance (2017b). Security guidance for critical areas of focus in cloud computing v4.0 [Online]. Available: https://downloads. cloudsecurityalliance.org/assets/research/security-guidance/security-guidance-v4-FINAL.pdf.

Cook, K.S., Cheshire, C., Rice, E.R., and Nakagawa, S. (2013). Social exchange Theory. In *Handbook of Social Psychology*, DeLamater, J. and Ward, A. (eds). Springer.

CrowdStrike, Inc. (2014). CSIX: CrowdStrike Intelligence Exchange [Online]. Available: https://www.crowdstrike.com/products/falcon-intelligence/.

CSIRT Gadgets. (2016). Bearded-avenger (CIF v3) [Online]. Available: http://csirtgadgets.org/bearded-avenger [Accessed February 2017].

Dalziel, H. (2014). *How to Define and Build an Effective Cyber Threat Intelligence Capability*. Syngress Publishing.

Danyliw, R., Meijer, J., and Demchenko, Y. (2007). The Incident Object Description Exchange Format (IODEF). *Internet Engineering Task Force (IETF), RFC-5070*.

Dshield (2017). [Online]. Available: https://www.dshield.org [Accessed January 2017].

Dunning, L.A. and Kresman, R. (2013). Privacy preserving data sharing with anonymous ID assignment. *IEEE Transactions on Information Forensics and Security*, 8(2), 402–413.

ElecticIQ (2017). EclecticIQ Platform [Online]. Available: https://www.eclecticiq.com/platform [Accessed January 2017].

Ellison, L. (2014). *Oracle Cloud and Your Datacenter Coexist*. Oracle Media Network.

ENISA: European Union Agency for Network and Information Security (2006). CERT cooperation and its further facilitation by relevant stakeholders.

ENISA: European Union Agency for Network and Information Security (2013). Detect, SHARE, Protect – Solutions for improving threat data exchange among CERTs.

ENISA: European Union Agency for Network and Information (2015). Cyber security information sharing: An overview of regulatory and non-regulatory approaches [Online]. Available: https://www.enisa.europa.eu/publications/cybersecurity-information-sharing/at_download/fullReport.

ENISA: European Union Agency for Network and Information (2017). ENISA threat landscape report 2016–15 top cyber threats and trends.

European CERTs/CSIRTs (2014). IntelMQ [Online]. Available: https://github.com/certtools/intelmq [Accessed April 2018].

Fadilpasic, S. (September 2016). Social media still an important tool for phishing. White paper, ITProPortal.

Falliere, N. and Chien, E. (2009). Zeus: King of the bots. Symantec Security Response. Available: https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/security-response-zeus-king-of-bots-09-en.pdf.

Ferrand, O. (2015). How to detect the cuckoo sandbox and to strengthen it? *Journal of Computer Virology and Hacking Techniques*, 11(1), 51–58.

FireEye Inc. (2012). Advanced targeted attacks – How to protect against the next generation of cyber attacks. Technical report.

FireEye Inc. (2014). Taking a lean-forward approach to combat today's cyber attacks. Technical report.

Fossi, M., Turner, D., Johnson, E., Mack, T., Adams, T., Blackbird, J., Entwisle, S., Graveland, B., McKinney, D., Mulcahy, J., and Wueest, C. (2010). Symantec global internet security threat report trends for 2009. White paper, symantec enterprise security, 15.

Furnell, S., Clarke, N., Beznosov, K., and Beznosova, O. (2007). On the imbalance of the security problem space and its expected consequences. *Information Management & Computer Security*, 15(5), 420–431.

Gerspacher, N. and Lemieux, F. (2010). A market-oriented explanation of the expansion of the role of Europol: Filling the demand for criminal intelligence through entrepreneurial initiatives. *International Police Cooperation: Emerging Issues, Theory and Practice*. Willan Publishing, Culompton.

Ghanaei, V., Iliopoulos, C.S., and Overill, R.E. (2016). Statistical approach towards malware classification and detection. *SAI Computing Conference*, IEEE, pp. 1093–1099.

Gilligan, J., Heitkamp, K., Dix, R., Palmer, C., Sorenson, J., Conway, T., Varley, W., Gagnon, G., Lentz, R., Venables, P., Paller, A., Lute, J.H., and Reeder, F. (2014). The economics of cybersecurity part II: Extending the cyber-security framework. Technical report, Armed Forces Communications and Electronics Association Cyber Committee.

Google Inc., Bengen, H., Metz, J., Buehlmann, S., and Alvarez Yara V.M. (2017). [Online]. Available: http://virustotal.github.io/yara [Accessed July 2017].

Guarnieri, C., Tanasi, A., Bremer, J., and Schloesser, M. (2016). Cuckoo sandbox [Online]. Available: https://www.cuckoosandbox.org.

Gundert, L. (2014). Producing a world-class threat intelligence capability. White paper, Recorded Future.

Heckman, K.E., Stech, F.J., Thomas, R.K., Schmoker, B., and Tsow, A.W., (2015). *Cyber Denial, Deception and Counter Deception*. Springer.

Heiser, J. and Nicolett, M. (2008). Assessing the security risks of Cloud computing [Online]. Available: https://www.gartner.com/doc/685308/assessing-security-risks-cloud-computing.

Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 50–57.

Holland R., Balaouras S., and Mak K. (2013). Five steps to build an effective threat intelligence capability. Forrester Research, Inc.

Hugh, P. (2016). What is threat intelligence? Definition and examples [Online]. Available: https://www.recordedfuture.com/threat-intelligence-definition.

Hutchins, E.M., Cloppert, M.J., and Amin, R.M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, 1, 80.

IBM (2017). X-Force exchange [Online]. Available: https://exchange.xforce.ibmcloud.com [Accessed January 2017].

iSightPartners (2014). What is cyber threat intelligence and why do I need it? iSIGHT Partners, Inc.

Issa, A. (2012). Anti-virtual machines and emulations. *Journal in Computer Virology*, 8(4), 141–149.

Johnson, C.S., Badger, M.L., Waltermire, D.A., Snyder, J., and Skorupka, C. (2016). Guide to cyber threat information sharing. Technical report, NIST Special Publication.

Kampanakis, P. (2014). Security automation and threat information-sharing options. *IEEE Security & Privacy*, 12(5), 42–51.

Keeling, C. (2013). Waking Shark II – Desktop cyber exercise – Report to participants. Technical report.

Khorshed, M.T., Ali, A.S., and Wasimi, S.A. (2012). A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. *Future Generation Computer Systems*, 28(6), 833–851.

Korstanje, M.E. (2016). Threat mitigation and detection of cyber warfare and terrorism activities. *Advances in Information Security, Privacy, and Ethics (AISPE) 2016*.

Landauer, T.K., Foltz, P.W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), 259–284.

Li, Z.-T., Lei, J., Wang, L., Li, D., and Ma, Y.-M. (2007). Towards identifying true threat from network security data. *Pacific-Asia Workshop on Intelligence and Security Informatics*, Springer.

Lightcyber (2016) Cloud based threat intelligence [Online]. Available: http://lightcyber. com/glossary/cloud-based-threat-intelligence [Accessed January2017].

Mandiant (2017) OpenIOC [Online]. Available: http://www.openioc.org [Accessed July 2017].

McMillan, R. (2013). *Definition: Threat Intelligence*. Gartner.

Miles, C., Lakhotia, A., LeDoux, C., Newsom, A., and Notani, V. (2014). VirusBattle: State-of-the-art malware analysis for better cyber threat intelligence. *2014 7th International Symposium on Resilient Control Systems (ISRCS)*, IEEE, pp. 1–6.

MITRE (2012). Cyber information-sharing models: An overview. Case no. 11-4486.

MITRE (2017a). ATT&CK: Adversarial Tactics, Techniques & Common Knowledge [Online]. Available: https://attack.mitre.org/wiki/Main_Page [Accessed July 2017].

MITRE (2017b). MAEC: Malware Attribute Enumeration and Characterization [Online]. Available: https://maec.mitre.org [Accessed July 2017].

MITRE (2017c). CAPEC: Common Attack Pattern Enumeration and Classification [Online]. Available: https://capec.mitre.org [Accessed July 2017].

MITRE (2017d). CRITS: Collaborative Research Into Threats [Online]. Available: https://crits.github.io/ [Accessed January 2017].

MITRE (2017e). Cyber Observable eXpression [Online]. Available: http://cyboxproject.github.io [Accessed July 2017].

Moriarty, K.M. (2011). Incident coordination. *IEEE Security & Privacy*, 9(6), 71–75.

Moriarty, K.M. (2012). Real-time inter-network defense (RID). *Internet Engineering Task Force (IETF), RFC-6545*.

Murdoch, S., and Leaver, N. (2015) Anonymity vs. trust in cyber-security collaboration. *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, ACM, pp. 27–29.

National High Tech Crime Unit of the Netherlands police, Europol's European Cybercrime Centre, Kaspersky Lab, Intel Security (2017). No more ransom project [Online]. Available: https://www.nomoreransom.org/index.html [Accessed July 2017].

Niddel Corp (2014). TIQ-Test – Threat Intelligence Quotient Test [Online]. Available: https://github.com/mlsecproject/tiq-test.

Oktavianto, D. and Muhardianto, I. (2013). *Cuckoo Malware Analysis*. Packt Publishing Ltd.

OrbITPeaple (2016). Migrating oracle databases to database cloud service.

Pace, C., Barysevich, A., Gundert, L., Liska, A., McDaniel, M., and Wetzel, J., (2018). *A Practical Guide for Security Teams to Unlocking the Power of Intelligence*. Cyber Edge Press.

Pepper, C. (2015). Applied threat intelligence. Technical report, Securosis.

Peretti, K. (2014). Cyber threat intelligence: To share or not to share – What are the real concerns? Privacy and security report, Bureau of National Affairs.

Phillips, G. (2014). Threatelligence v0.1 [Online]. Available: https://github.com/syphon1c/Threatelligence.

Pinto, A. and Sieira, A. (2015). Data-driven threat intelligence: Useful methods and measurements for handling indicators. *27th Annual FIRST Conference*, June.

Piper, S. (2013). *Definitive Guide to Next Generation Threat Protection*. CyberEdge Group, LLC.

Ponemon (2013). Live threat intelligence impact report 2013. Technical report.

Ponemon (2014). Exchanging cyber threat intelligence: There has to be a better way. Technical report.

Ponemon (2015). Second annual study on exchanging cyber threat intelligence: There has to be a better way. Technical report.

Poputa-Clean, P. and Stingley, M. (2015). Automated defense – Using threat intelligence to augment security [Online]. *SANS Institute InfoSec Reading Room*. Available: https://www.sans.org/reading-room/whitepapers/threats/automated-defense-threat-intelligence-augment-35692.

Ray, J. (2015). *Understanding the Threat Landscape: Indicators of Compromise (IOCs)*. Verisign.

RecordedFuture (2017). Threat intelligence, information, and data: What is the difference? [Online]. Available: https://www.recordedfuture.com/threat-intelligence-data/.

Richards, K. (2009). *The Australian Business Assessment of Computer User Security (ABACUS): A National Survey*. Australian Institute of Criminology.

Rieck, K. (2013). Malheur [Online]. Available: http://www.mlsec.org/malheur.

Rieck, K., Trinius, P., Willems, C., and Holz, T. (2011). Automatic analysis of malware behavior using machine learning. *Journal of Computer Security*, 19(4), 639–668.

Ring, T. (2014). Threat intelligence: Why people don't share. *Computer Fraud & Security*, 2014(3), 5–9.

Rouse, M. (2015). Hybrid Cloud [Online]. TechTarget.

Sauerwein, C., Sillaber, C., Mussmann, A., and Breu, R. (2017). Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. *Towards Thought Leadership in Digital Transformation: 13. Internationale Tagung Wirtschaftsinformatik* [Online]. Available: http://aisel.aisnet.org/wi2017/track08/paper/3.

Schneier, B. (2000). Software complexity and security [Online]. Crypto-Gram.

Seredynski, M., Bouvry, P., and Klopotek, M.A. (2007). Modelling the evolution of cooperative behavior in ad hoc networks using a game based model. *IEEE Symposium on Computational Intelligence and Games, 2007*, pp. 96–103.

Shackleford, D. (2015). Who's using cyberthreat intelligence and how? – A SANS Survey. Technical report, SANS Insitute.

Shackleford, D. (2016). The SANS state of cyber threat intelligence survey: CTI important and maturing. Technical report, SANS Institute.

Sillaber, C., Sauerwein, C., Mussmann, A., and Breu, R. (2016). Data quality challenges and future research directions in threat intelligence sharing practice. *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, ACM, pp. 65–70.

Skopik, F., Settanni, G., and Fiedler, R. (2016). A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security*, 60, 154–176.

Soltra (2017). Soltra Edge [Online]. Available: http://www.soltra.com/en [Accessed January 2017].

Steele, R.D. (2007a). Open source intelligence. In *Handbook of Intelligence Studies*. Johnson, L.K. (ed.), Routledge.

Steele, R.D. (2007b). Open source intelligence. In *Strategic Intelligence.* Johnson, L.K. (ed.). Praeger.

Steele, R.D. (2014). Applied collective intelligence: Human-centric holistic analytics, true cost economics, and open source everything. *Spanda Journal*, 2, 127–137.

Strom, B. (2016). ATT&CK Gaining Ground [Online]. Available: https://www.mitre.org/capabilities/cybersecurity/overview/cybersecurity-blog/ attck%E2%84%A2-gaining-ground.

Symantec. (2016). Internet security threat report. Technical report.

The Snort team (2017). Snort [Online]. Available: https://www.snort.org [Accessed July 2017].

Threatconnect (2017). Threatconnect platform [Online]. Available: https://www.threatconnect.com/platform [Accessed January 2017].

ThreatQuotient (2017). THREATQ [Online]. Available: https://www.threatq.com/threatq [Accessed January 2017].

Tounsi, W. and Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber-attacks. *Computers & Security*, 72, 212–233 [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0167404817301839.

Tounsi, W., Cuppens-Boulahia, N., Cuppens, F., and Garcia-Alfaro, J. (2013). Fine-grained privacy control for the RFID middleware of EPCGlobal networks. *Proceedings of the Fifth International Conference on Management of Emergent Digital EcoSystems*, ACM, pp. 60–67.

Trost, R. (2014). Threat intelligence library: A new revolutionary technology to enhance the SOC battle rhythm! Briefing, Blackhat-Webcast.

Verizon (2015). Data breach investigations report. Technical report.

Verizon (2017). Vocabulary for event recording and incident sharing [Online]. Available: http://veriscommunity.net [Accessed July 2017].

Wagner, C., Dulaunoy, A., Wagener, G., and Iklody, A. (2016). MISP: The design and implementation of a collaborative threat intelligence sharing platform. *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, ACM, pp. 49–56.

Williamson, W. (2016). Distinguishing threat intelligence from threat data [Online]. *Security Week*. Available: http://www.securityweek.com/ distinguishing-threat-intelligence-threat-data.

Yamakawa, N. (2014). Threat intelligence climbs the ladder of enterprise proliferation. Technical report, 451 Research.

Zheng, D.E. and Lewis, J.A. (2015). Cyber threat information sharing – recommendations for congress and the administration. Report Center for Strategic and International Studies (CSIS), Strategic Technologies Program.

Zurkus, K. (2015). Threat intelligence needs to grow up. White paper, CSOfromIDG.