Application of Artificial Intelligence and Machine Learning in Producing Actionable Cyber Threat Intelligence

Reza Montasari*, Fiona Carroll, Stuart Macdonald, Hamid Jahankhani, Amin Hosseinian-Far, and Alireza Daneshkhah

Hillary Rodham Clinton School of Law, Swansea University,
Richard Price Building, Sketty Ln, Sketty, Swansea, SA2 8PP, United Kingdom.
{montasarireza@gmail.com}
http://www.swansea.ac.uk
Cardiff School of Technology, Cardiff Metropolitan University,
Llandaff Campus, Western Avenue, Cardiff, CF5 2YB, United Kingdom.
{FCarroll@cardiffmet.ac.uk}
http://www.cardiffmet.ac.uk
Hillary Rodham Clinton School of Law, Swansea University,
Richard Price Building, Sketty Ln, Sketty, Swansea, SA2 8PP, United Kingdom.
{s.macdonald@swansea.ac.uk}
http://www.swansea.ac.uk
Information Security and Cyber Criminology, Northumbria University,
110 Middlesex Street, London, E1 7HT, United Kingdom.
{hamid.jahankhani@northumbria.ac.uk}
http://www.london.northumbria.ac.uk
Faculty of Business and Law, University of Northampton,
Waterside Campus, University Drive, Northampton, NN1 5PH, United Kingdom.
{Amin.Hosseinian-Far@northampton.ac.uk}
https://www.northampton.ac.uk/
School of Computing, Electronics and Mathematics, Coventry University,
Priory Street, Coventry, CV1 5FB, United Kingdom.
{ac5916@coventry.ac.uk}
https://www.coventry.ac.uk/

Abstract. Cyber Threat Intelligence (CTI) can be used by organisations to assist their security teams in safeguarding their networks against cyber-attacks. This can be achieved by including threat data feeds into their networks or systems. However, despite being an effective Cyber Security (CS) tool, many organisations do not sufficiently utilise CTI. This is due to a number of reasons such as not fully understanding how to manage a daily flood of data filled with extraneous information across their security systems. This adds an additional layer of complexity to the tasks performed by their security teams who might not have the appropriate tools or sufficient skills to determine what information to prioritise and what information to disregard. Therefore, to help address the stated issue, this paper aims firstly to provide an in-depth understanding of what

*

CTI is and how it can benefit organisations, and secondly to deliver a brief analysis of the application of Artificial Intelligence and Machine Learning in generating actionable CTI. The key contribution of this paper is that it assists organisations in better understanding their approach to CTI, which in turn will enable them to make informed decisions in relation to CTI.

Keywords: Cyber security, Threat intelligence, Artificial intelligence, Machine learning, Cyber physical systems, Digital forensics, Big data

1 Introduction

Cyber threats are constantly growing in frequency and complexity [19] [17] [18] [20]. Through the use of intrusion kill chains, campaigns and customised tactics, techniques and procedures, cyber criminals are able to bypass organisations' security controls [23] [22] [24]. Cyber Security (CS) breaches and outages have been widely covered in the media, and statistics concerning the number of cyber-attacks are available in a variety of sources [6] [16] [21] [25]. However, despite many CS breaches, there is little expert analysis of the areas that organisations should prioritise in order to increase their effectiveness in addressing known threats while also minimising the risk from evolving attacks [28]. One of the ways to help mitigate security breaches is by developing and implementing robust CTI. CTI is focused on analysing trends and technical developments in three areas of CS, Hacktivism and Cyber Espionage. CTI is used by nations states as an efficient solution to devise preventive CS measures in advance and as a result to uphold international security.

CTI is a branch of CS that concerns the contextual information surrounding cyber-attacks, i.e. the understanding of the past, present, and future tactics, techniques and procedures (TTPs) of a wide variety of threat actors. It is actionable and timely and has business values in that it can inform the security teams in organisations of adversarial entities so that they can prevent them. CTI is also a proactive security measure that involves the gathering, collation and analysis of information concerning potential attacks in real time so as to prevent data breaches and subsequent adverse consequences. Its primary objective is to deliver detailed information on the security threats that pose a higher risk to an organisation's infrastructure and simultaneously guide the security teams on preventative actions.

By providing continuously updated threat data feeds, CTI can enable security teams to defend against cyber-attacks before they can enter their networks or detect already malicious activities on enterprise networks. For instance, CTI can assist the teams in gaining a detailed understanding of the adversary and their modus operandi. This, in turn, enables them to improve their protection against specific attack methods known to be used by the adversary, and helps produce actionable information that can enable decision makers to comprehend

their operational risks and better prioritise and allocate resources. Therefore, to be effective, CTI must be able to provide context and to be understood by decision makers. While CTI's main focus is on traditional IT systems, industrial control system (ICS) and network operators could also benefit from this capability given that many of the threats to ICS are facilitated by traditional IT networks. A CTI network can be considered as a combination of regular updating and learning feeds that develop the basis of powerful layered network security. Such threat feeds enable individual devices and networks to take advantage of the intelligence of numerous devices to safeguard their endpoints and networks.

Considering the above, many organisations attempt to include threat data feeds into their networks or systems without fully understanding how to deal with a daily flood of data filled with extraneous information across their security systems. This adds an additional layer of complexity to the tasks performed by security analysts who might not have the appropriate tools to determine what information to prioritise and what information to disregard. Therefore, to address the stated issues, this paper aims firstly to provide an in-depth understanding of what CTI is and how it can benefit organisations, and secondly to analyse the application of Artificial Intelligence (AI) and Machine Learning in generating actionable CTI. The key contribution of this paper is that it assists organisations in better understanding their approach to CTI, which in turn will enable them to make informed decisions in relation to CTI.

The remainder of this paper is structured as follow: Section 2 provides a brief overview of CTI and its benefits. Section 3 discusses phases of our recommended six-phase CTI Cycle (CTIC) and how each phase can be utilised to provide intelligence, help to guide decisions, shorten the information aggregation and dissemination timelines, and assist organisations in protecting their networks from cyber-attacks. Section 4 analyses the application of Artificial Intelligence (AI) and Machine Learning (ML) in producing actionable CTI. In Section 5, a discussion is provided, and finally the paper is concluded in Section 6.

2 Cyber Threat Intelligence

2.1 Overview of CTI

CTI is an ambiguous concept with numerous definitions attributed to it that are based on different procedural viewpoints and competitive imperatives. One definition that provides a comprehensive description is provided by McMillan [12], who defines CTI as:

> *evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard.*

Despite its ambiguity, CTI should have three main characteristics including, (1) evidence based: cyber threat evidence may be acquired from malware analysis to ensure that the threat is valid, (2) utility: there must have some utility for organisations to have a positive impact on security incidents, and (3) actionable: the gathered CTI must drive not only data or information but also security control action [10]. It must include the combination of information detailing possible threats with a solid insight into network structure, operations, and activities. In order to produce this evidence-based knowledge, information on the mechanisms and indicators, i.e. threat feeds, will need to be put into context by contrasting it with the core knowledge of network activity. The process of gathering and collation of threat feeds will result in threat intelligence, "which then informs 'security analytics' to improve chances of detection" [4]. Security analytics in a network defence environment often consists of one of the following two forms, both of which are informed by CTI: 'Big data' platform processing large amounts of network data to determine trends, and 'Security information and event management (SIEM) infrastructure' to automate the detection of anomalous activities.

CTI is collected by continuously analysing large quantities of threat data with the aim of organising and adding context to cyber threat activities, trends and attacks. It can be derived from external threat feeds, internal networks, analysis of historical attacks, and research. For instance, it can be generated through the aggregation of fused, heterogeneous and highly reliable sources of data such as security networks, web crawlers, botnet monitoring service, spam traps, research teams, the open web, dark web, deep web, social media, and collected historical data about malicious objects. All the aggregated data is then carefully examined and processed in its entirety (often in real-time) through several pre-processing techniques, including statistical criteria, expert systems (such as sandboxes, heuristics engines, similarity tools, behaviour profiling etc.), security analysts' validation and whitelisting verification.

2.2 Types of Threat Intelligence

CTI can be classified into four main types as depicted in Figure 1 in relation to information assortment, knowledge analysis and intelligence consumption. These consist of Tactical, Technical, Operational and Statistical threat intelligence [3]. The followings describe each type.

Tactical Cyber Threat Intelligence Tactical CTI (TaCTI) focuses on the techniques and procedures of threat actors such as methodologies, tools, and tactics, relies on sufficient resources and includes certain specific measures against malicious actors attempting to infiltrate a network or system. TaCTI should be used to evaluate real-time events, investigations, and activities, and to provide support for day-to-day operations and events such as the development of signatures and indicators of compromise (IOCs). It must be aimed at the immediate future and identifies simple IOCs (such as malicious IP addresses, URLs,

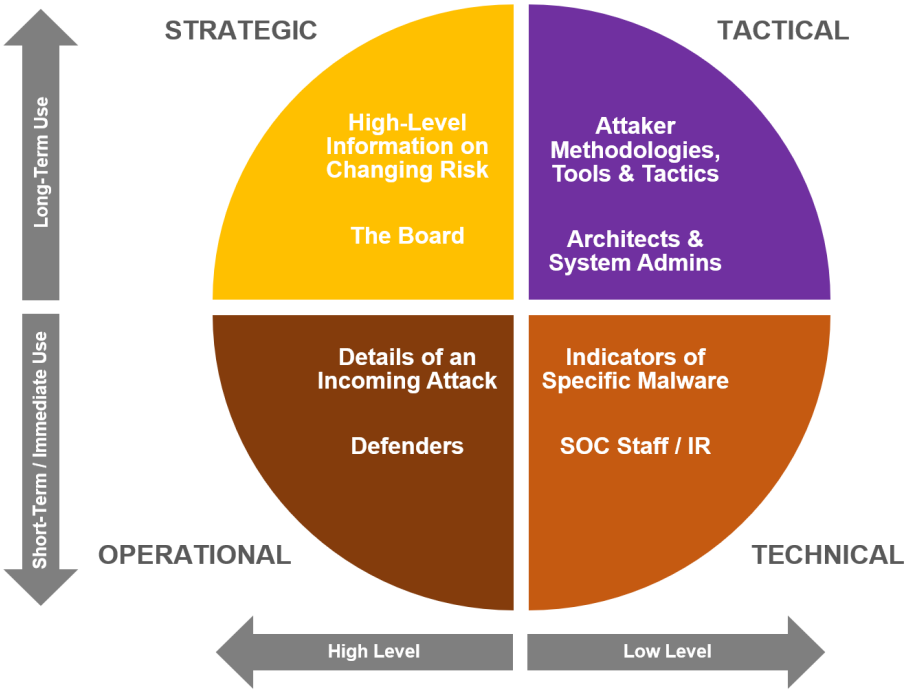

Fig. 1. Types of threat intelligence.

file hashes and known malicious domain names). If implemented properly, it can provide organisations with a deeper understanding of how they have been previously attacked and how they can mitigate such attacks. TaCTI is often automated and machine-readable enabling security products to ingest it through feeds or API integration. It is considered to be the easiest type of intelligence to be produced, and as a result, it can be found through open source and free feeds. It must be noted that TaCTI has a short lifespan given that IOCs can become outdated in a short period of time.

Technical Cyber Threat Intelligence Technical CTI (TeCTI) should focus on the technical clues that are indicative of a CS threat such as the subject lines to phishing emails, fraudulent URLs or specific malware. TeCTI enables security analysts to determine what to look for, rendering it valuable for analysing social engineering attacks. However, in the financial sector such as the banking sector, penetration testing no longer appears to be sufficient to shield sensitive business sectors. Considering this, the UK Financial Authorities have recommended several steps which can be found in [2] to protect financial institutions from cyber threats.

Operational Cyber Threat Intelligence Operational CTI (OCTI) pertains to details of specific events associated with the cyberattack in order to facilitate an understanding of the nature, severity, timing, and intent of specific attacks. OCTI involves cybersecurity professionals learning about threat actors and is focused on addressing the 'attribution' elements of CTI, such as 'who', 'why' and 'how' questions. In this context, 'who' refers to threat actors, 'why' addresses the motivation or intent, and 'how' consists of tactics, techniques and procedures (TTPs) that adversaries use to carry out attacks. The attribution elements offer context, and context, in turn, provides insight into how attackers plan, conduct, and sustain campaigns and operations. Such an insight is considered to be operational intelligence which cannot be produced by machines alone. If implemented properly, OCTI will be able to provide highly specialised and technically focused intelligence to guide and assist with the response operations.

Thus, OCTI should be based on details of the specific incoming attack and evaluation of an organisation's capability in determining future cyber-threats. It must be able to assess specific attacks associated with events, investigations and malicious behaviour, and provide an understanding that can guide and support response to specific incidents. This type of CTI requires Cyber Security Analysts who can convert data into a format that is readily usable by end-users. Despite the fact that OCTI necessitates more resources than that required by TaCTI, it offers a longer valuable lifespan. This is due to the fact that attackers will not be able to alter their TTPs in the same way that they could easily change their tools. OCTI is often most beneficial for those cybersecurity specialists operating in security operations centers (SOCs) who are in charge of conducting routine operations. Professionals operating in CS branches such as Vulnerability Management, Incident Response and Threat Monitoring are the main customers of OCTI as it assists them with becoming more capable and effective at their assigned tasks [3] [32].

Strategic Cyber Threat Intelligence Strategic CTI (SCTI) must be aimed at long-term issues and be based on high-level information on CS modus operandi, threats, details concerning impact of fund on different cyber activities, attack tendencies, and the effect of high-level business assortments. Therefore, SCTI must be employed (1) to evaluate disparate pieces of information to establish unified views, and (2) to develop an overall picture of the intent and capabilities of cyber threats (such as the actors, tools and TTPs) through the identification of trends, patterns, and evolving threats with a view to inform decision makers. An effective SCTI should also be able to enable time alerts of threats against organisations' important assets such as IT infrastructure, employees, customers, and applications. This information should be in the format of reports, whitepapers, policy documents, or publications in the industry and must then be presented to high-level executives, such as Chief Information Security Officers (CISO) for the purposes of decision making.

Furthermore, SCTI can be used as a means to understand how global events, foreign policies, and other long-term national and international movements can influence the CS of an organisation. This understanding can assist decision-makers in understanding cyber threats against their organisations more effectively. In turn, this knowledge can enable them to make CS investments that safeguard their organisations and are aligned with its strategic priorities [32]. SCTI is the most challenging type of intelligence to produce as it entails human collection and analysis that require an in-depth knowledge of both CS and global geopolitical situation. To this end, often, senior leadership is required to perform critical evaluations of cyber threats against their organisations.

2.3 Benefits of Cyber Threat Intelligence

If implemented effectively, CTI provides substantial benefits as threat information can be shared in machine-readable formats that can be promptly obtained and imported for immediate use by security incident and event management (SIEM) tools and CTI platforms (CTIPs). CTI can enable the development of a focused defence against specific threats as well as the insight to apply the appropriate CS tools and solutions to protect organisations. Furthermore, CTI can provide organisations with context such as intelligence about the attackers, their motivation and capabilities and indicators of compromise (IoCs) in their system to investigate. This information will enable organisations to make informed decisions about their security. Based on its classification, described in the previous section, CTI offers four types of tactical, technical, operational, and strategic benefits as shown in Table 1.

In addition to the above, CTI can contextualise threat information that is more meaningful for the end-user. This, in turn, reduces ambiguity, enhances situational awareness, and results in more informed risk management and security investment. Furthermore, CTI can assist vulnerability management teams in prioritising the most vital susceptibilities more accurately with access to the external understandings enabled by CTI. Similarly, comprehending the existing threat landscape (comprising key insights on threat actors and their modi operandi) that CTI provides can augment other high-level security processes such as fraud prevention and risk analysis. As well as assisting organisations to protect their networks, CTI can also enable them to regulate costs of sustaining their network security and provide the security teams with the knowledge they require to concentrate on what really matters.

3 Cyber Threat Intelligence Cycle

To produce intelligence (the final product of the CTI cycle), organisations would firstly require to collect raw data. This raw data represents simple facts that are available in large quantities such as IP addresses or logs. On its own, the raw data has limited usefulness until it is converted to information through

Table 1. Benefits of Tactical, Technical, Operational and Strategic CTI.

Types of CTI	Benefits
Tactical CTI	• enables organisations to develop a proactive cybersecurity posture and to strengthen overall risk management policies. • informs better decision-making during and after the detection of a cyber-attack. • assists with a cybersecurity posture that is predictive. • facilitates enhanced detection of advanced threats.
Technical CTI	• connects details associated with attacks rapidly and accurately. • provides rapid response to new indicators. • enables security analysts to determine what to look for rendering it valuable for analysing social engineering attacks.
Operational CTI	• provides context and relevance to a large amount of data that enable organisations to gain better insight into how threat actors plan, carry out, and sustain offensives and major operations. • enables organisations to detect and respond to cyber-attacks more swiftly and assisting them in preventing future incidents.
Strategic CTI	• provides a more in-depth situational awareness. • assists decision-makers in understanding the risks posed by cyber threats to their organisations. • enables decision-makers in making cybersecurity investments that effectively defend their organisations and are aligned with its strategic priorities. • produces an organisational situational awareness that will help existing and future security strategies.

data processing for the purposes of producing a valuable output. An example of information is a collated series of logs that display an increase in suspicious activities. Intelligence can then be produced by processing and analysing this information, which must be able to inform decision making. As an example, the collated data is placed in context along with prior incident reports in relation to similar activities that enable the development of a strategy to reduce cyber-attacks [5]. Figure 2 represents a useful model that visualises the processing of raw data into a complete intelligence product.

The Cyber Threat Intelligence Cycle (CTIC), that produces intelligence, must be a methodical, continuous process of analysing potential threats to detect a suspicious set of activities that can threaten organisations' systems, networks, information, employees, or customers. It must visually represent and evaluate a number of specific intrusion sensor inputs and open source information to determine specific threat courses of action [11]. Therefore, the CTIC should be a process whereby raw data and information are identified, collected and then built into a complete intelligence for use by decision makers. The model must also be able to support organisations' risk management strategies and the information security teams' decision-making. To exploit the benefits of CTI, it is essential to define both appropriate objectives as well as relevant use cases. Since the CTIC

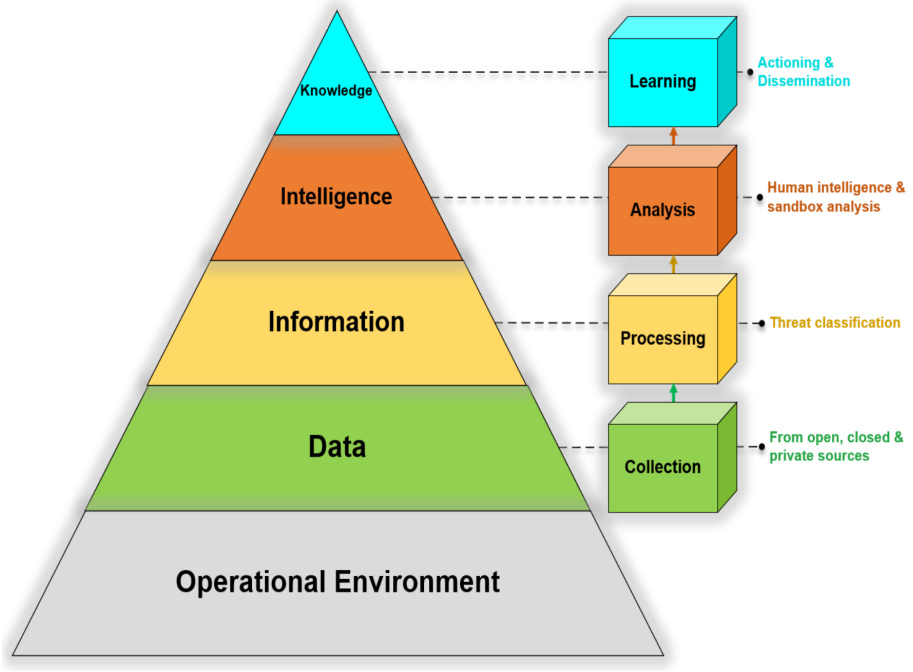

Fig. 2. Processing of raw data into a complete intelligence product.

is intended to produce intelligence, the information security teams must be able to formulate new questions and identify gaps in knowledge during this lifecycle. In turn, this should result in the requirements development. Furthermore, in order for a CTIC to be an effective intelligence scheme, it must be based on iterative phases that can become more sophisticated over time.

Therefore, considering the discussion above, we recommend a six-phase cycle consisting of the following stages: Planning and Direction, Data Collection, Data Processing and Exploitation, Data Analysis and Production, Dissemination and Integration, and Feedback. All steps in the cycle must also incorporate an Evaluation process and a Review process that must be performed simultaneously throughout the entire six phases so as to ensure that the necessary materials are being processed accurately and that the original questions are being addressed effectively. Figure 3. represents our recommended CTIC along with the description of each phase.

Fig. 3. Six-phase cyber threat intelligence cycle.

3.1 Planning and Direction

The CTI's production life cycle starts with requirements or questions unique to the end-user that should be answered. After the CTI requirements have been identified and prioritised, a data collection plan comprising identification and evaluation of information sources should be created. Planning and Direction is the first phase of the CTI, that is intended to produce actionable threat intelligence based on a set of accurate questions that should enable the development of actionable threat intelligence. These questions must focus on a single fact, event, or activity as opposed to broad, open-ended questions [8]. A key aspect of this phase should be understanding who will consume and benefit from the complete product. Next, individuals involved with planning and direction should be able to establish the precise requirements of the consumer, called intelligence

requirements (IRs), and prioritise intelligence requirements (PIRs). These IRs and PIRs must be based on certain factors such as how closely they comply with organisations' core values and must determine what data and information are required and how it should be collected. This output is often systematised in an intelligence collection plan (ICP) [5]. It is important that this phase involves substantial interaction between the consumer and producer.

3.2 Collection

The next step in the CTI is the Collection phase, that involves gathering raw data [26] [15] [14] [13]. This data must be meaningful to the organisation and able to address the initial CTI requirements established in the first phase. Raw data can be gathered from a wide variety of sources such as internal ones including network event logs and records of past incident responses and external ones from the open web, the dark web, and technical sources [8]. The data Collection phase must be timely and accurate, as well as being applicable to deal with incidents that can occur or are occurring. Understanding which sources are likely to generate the desired information, be reliable, and provide information that can be used in a timely manner is a complex process that necessitates thoughtful and robust planning and direction to assist in isolating the signals from the noise.

Instances of CTI data sources consist of traditional Security Information and Event Management (SIEM) tools (such as network monitors, firewalls, intrusion detection systems), dedicated CTI data feeds, vulnerability and malware databases, and the system users. It is through these data sources that indicators of compromise (IoCs) can be identified, documented, and further analysed. IoCs which represent threat data concerns measurable events that can be classified as either network-based or host-based events. Examples of network-based IOCs comprise email addresses, subject line and attachments, connections to specific IP addresses or web sites, file hashes, and fully qualified domain names utilised for botnet command and control server connections. Instances of host-based IoCs consist of the presence of filenames on a local drive, programs and processes that are running on a machine, and creation or modification of dynamic link libraries (DLLs) and registry keys [31]. Furthermore, IoCs can also include vulnerability information, such as the personally identifiable information of customers, raw code from paste sites, and text from news sources or social media.

3.3 Processing and Exploitation

Processing and Exploitation is the third phase in the CTCI, that involves converting the raw data into intelligence. The raw data that have been collected from multiple data sources must be integrated and sorted in order to produce more consistent, accurate, and useful information than that provided by any individual data source. To achieve this, one needs to sort and fuse it with other data sources by organising it with metadata tags and filtering out redundant information or false positives and negatives [8]. During this phase, both human

and machine capabilities are needed to address the IRs for the engagement while complying with the tenets of intelligence. Given that data is collected from millions of log events and indicators every day, processing such data manually is extremely cumbersome. Thus, collecting data must be automated in order to extract meaningful intelligence from it. One of the best ways to achieve this is to deploy solutions such as SIEM since it facilitates structuring and correlating event data with rules that can be established for various use cases (even though it can only deal with a limited number of data types). See section 4 for details on more powerful data processing solutions.

3.4 Analysis and Production

Analysis and Production is the next phase in the CTIC, where analysts will need to make sense of the processed data. The objective of this phase is to look for possible security threats and inform the relevant audience in a format that achieves the intelligence requirements defined in the Planning and Direction phase [8]. The analysis must be determined based on three elements of actors, intent, and capability, with consideration given to their tactics, techniques, and procedures (TTPs), motivations, and access to the intended targets. By examining these three elements, it is often possible to make informed, forward-leaning strategic, operational, and tactical assessments. Furthermore, during this phase, analysts must be able to produce intelligence products, i.e. the answers to the questions posed earlier during the requirements gathering, and identify connection between the technical indicators, attackers, their motivations and aims, and information related to the target [30]. This should then result in informative and proactive decision-making. To do so, analysts will need to employ a wide range of quantitative and qualitative analytical techniques to evaluate the significance and implications of processed information, merge contrasting items of information to find patterns, and then interpret the meaning of any newly developed knowledge.

Additionally, they will need to apply a variety of approaches to assess the reliability of the sources and the material collected and to ensure accurate and unbiased evaluations that need to be predictive and actionable. It is also vital for any potential ambiguities to be handled properly, for instance, by determining how the questions have been addressed. Analysis phase must be accurately documented and efficiently implemented to assist organisations in utilising the collected data more effectively. This should be followed by a timely dissemination of intelligence to internal and external audiences in a format understandable to them such as threat lists and peer-reviewed reports.

3.5 Dissemination and Integration

Dissemination phase should involve communicating and distributing the complete product in a suitable form to its intended consumers. In order for CTI to be actionable, it must be delivered to the right audience at the right time, i.e. the occurrence of dissemination should correspond to the time period on which the

content is based. For instance, operational material requires to be regularly conveyed whilst strategic content will be more sporadic. The Dissemination phase must also be traceable in order that there is continuation between one CTIC and the next and that the learning is not lost. One of the ways in which this can be achieved is by utilising ticketing systems that integrate with the consumers' other security systems to trace each stage of the CTIC. Everytime a new intelligence request is made, tickets can be submitted, written up, reviewed, and fulfilled by different audience in one place. By obtaining feedback and refining existing IRs or creating new ones, the CTI cycle can commence again [8].

3.6 Feedback

The Feedback is the final stage in CTIC, in which a complete intelligence has been developed linking it to the original Planning and Direction phase. During this phase, individual/s who made the original request reviews the complete intelligence product to establish whether their questions have been addressed. This assists in informing the objectives and procedures of the next CTI cycle, once again highlighting the importance of documentation and continuation.

4 Application of Artificial Intelligence and Machine Learning in Producing Actionable CTI

AI and ML are two promising fields of research that can significantly improve CS measures. For instance, CS applications using AI and ML can perform anomaly detection on a network more effectively than those performed by traditional methods. With rapid pace of development and the desire for more effective countermeasures, AI and ML come as a natural solution to the problem of coping with the ever-growing number of cyber-attacks. This interdisciplinary endeavour has created a joint link between computer specialists and network engineers in designing, simulating and developing network penetration patterns and their characteristics. Some of these diverse methods are directed towards: Multi-Agent Systems of Intelligent Agents, Neural Networks, Artificial Immune Systems and Genetic Algorithms, Machine Learning Systems, including: Associative methods, Inductive Logic Programming, Bayes Classification, Pattern Recognition Algorithms, Expert Systems, and Fuzzy Logic.

Examples of AI and ML applications that can be used in CS solutions include: Spam Filter Applications, Network Intrusion Detection and Prevention, Fraud Detection, Credit Scoring and Next-Best Offers, Botnet Detection, Secure User Authentication, Cyber Security Ratings, and Hacking Incident Forecasting, etc. For instance, by determining certain distinctive features, AI and ML systems can be trained to analyse and distinguish between a normal software and malware. These features can comprise: accessed APIs, accessed fields on the disk, accessed environmental products, consumed processor power, consumed bandwidth, and

amount of data transmitted over the internet. By utilising these distinct features, the system is developed. Once a test software is fed to the system, it can then determine whether the software is a malware or not by analysing these distinct features [27].

In the specific context of CTI, organisations can utilise AI and ML methods to automate data acquisition and processing, combine with their existing security solutions, absorb unstructured data from disjunctive sources, and then link information from different places by adding context on compromise and modi operandi of malicious actors. This is particularly important in the context of Big Data, due to the scales of which its processing necessitates automation to be comprehensive. This processing should comprise the fusion of data points from a wide range of sources such as open web, deep web, dark web, and technical sources in order to draw up the most robust strategy. This can help to convert these large quantities of data into actionable CTI. Furthermore, by means of AI and ML techniques, data can be structured into categories of entities based on their names, properties, relationships to each other, and events by separating concepts and assembling them together. This will facilitate robust searches on the categories, enabling the automation of data sorting as opposed to sorting data manually [29]. In addition, AI and ML techniques can be applied for the purposes of structuring text in many languages through Natural Language Processing (NLP). For instance, NLP can be exploited to analyse text from almost infinite unstructured documents across a wide range of languages and categorise them by means of language-independent groups and events [8].

Moreover, ML techniques can be developed to categorise text into groups prose, data logs, or code, and remove ambiguities between entities with the same name through the use of contextual clues in the surrounding text. ML and statistical methodology can be implemented to sort entities and events even further based on significance, for instance by evaluating risk scores to malicious entities. Risk scores can be calculated by the ML trained on an already examined dataset. Classifiers such as risk scores deliver both a judgment and context describing the score since different sources verify that this IP address is malicious. Automating risk classification saves substantial time by sorting through false positives and determining what to prioritise. In addition, ML can be used to predict events and entity properties by producing predictive analysis models more accurately than those created by humans based on deep pools of data that have been previously mined and categorised [8] [29]. It is also likely that ML techniques could function as active sensors that feed data into a common threat intelligence network that can be employed by the entire user base. The above said, the process of applying ML and AI methods at the different levels of CTI is at very different stages. For instance, studies in Operational Intelligence type are still in the experiment and research stage and as a result necessitate substantial resources.

5 Discussion

Cyber threats are constantly growing in frequency and complexity, and the threat landscape is continually evolving. Through the use of various customised TTPs, cyber criminals are able to bypass organisations' security controls. As a result, organisations are under constant pressure to manage security vulnerabilities. One of the ways to help address security vulnerabilities is by developing and implementing robust CTI. CTI is based on traditional intelligence gathering and processing activities used to track, analyse and counter CS threats. The information collected through CTI can enable the security teams to identify, prepare, and impede cyber-attacks that can pose risk to the data integrity. CTI feeds can assist organisations in this process by identifying common IoCs and suggesting required steps to stop cyber-attacks. The most common IoCs consist of [7]:

- IP addresses, URLs and Domain names: An example is malware that targets an internal host that is communicating with a known threat actor.
- Email addresses, email subject, links and attachments: An example is a phishing attempt that depends on a user clicking on a link or attachment and starting a malicious command.
- Registry keys, filenames and file hashes and DLLs: An example is an attack from an external host which has already been flagged or that is already infected.

Robust CTI feeds could potentially have millions of computers functioning as security sensors which feed CTI to the entire users subscribing to that feed. At the same time, millions of security updates can automatically and seamlessly take place on the daily basis to end users and networks.

It is important to note that in order for organisations to be able to access CTI when needed, they will need to incorporate it into their broader security model as an essential component that enhances every other function (as opposed to a separate function). Incorporating CTI into security solutions that organisations already employ reinforces their security postures. Such an integration can enable security operations teams to respond to and process the alerts more effectively by helping automatically to prioritise and sieve through security threats. It is also imperative that there will be a clear distinction between threat data and threat intelligence. Without intelligence, data will not be able to provide the predictive knowledge required to detect threats before they can enter organisations' networks.

6 Conclusion

CTI can add significant values to organisations' security functions as well as to every level of government entity such as Chief Information Security Officers (CISOs), police chiefs, policy makers, information technology specialists, law

enforcement officers, security officers, accountants, and terrorism and criminal analysts. If implemented properly, CTI can facilitate better understanding into cyber threats, enabling a faster, more targeted response and resource development and allocation [9]. It can enable decision makers to define acceptable business risks, create controls and budgets, make equipment and staffing decisions, provide insights that guide and support incident response and post-incident activities (operational/technical intelligence), and advance the use of indicators by validating, prioritising, specifying the length of time an indicator is valid for (tactical intelligence). Likewise, when timely, relevant, and actionable, CTI can enable organisations to operate more efficiently and effectively by gaining the advantage they require to combat cyber-attacks prior to loss being incurred. Furthermore, by utilising CTI, organisations will be able to update their endpoint and network security proactively in real-time without the need to update their network security environments manually. For instance, in cases where one endpoint device faces a threat, that intelligence will be able to update the larger CTI network automatically. This enables organisations to stay ahead of cyber threats and attackers consistently and ensure that they are safeguarded against the latest cyber-attacks.

As security vendors compete with each other to deal with the consumer demand for assistance with the increasing number of threats, the market is now providing a wide range of CTI tools. However, not all tools are developed equal. For a successful implementation of security at this level to function effectively, the tool must be able to search through the vast and miscellaneous stretch of online content for potential security threats at every second. Therefore, a CTI security solution must be customizable and capable of providing clear and complete investigation with advanced analytics such as AI and ML that can be adapted to specific behavioural activities [7].

It is envisioned that over the next few years the inclusion of CTI into organisations' and governments' operations will become increasingly vital, as all levels and employees are forced to respond to the cyber threats. It is also envisaged that in the near future, cloud-based network security and secure web gateways fed by threat intelligence replace legacy firewalls, appliances, software and much of the resources required to patch and update in traditional environments [1]. As a future research direction, one area of CTI that has remained underexplored concerns the application of Multi-Agent Systems (MASs) in Tactical Cyber Threat Intelligence (TCTI). Therefore, experiments should be performed with the application of MASs to determine whether it can be an appropriate method for the needs of the CTI.

References

1. Avast. What is Threat Intelligence?. `https://smb.avast.com/answers/threat-intelligence.`, 2020. Avast. (Accessed: 07-03-2020).

2. CBEST. CBEST Intelligence-Led Testing: CBEST Implementation Guide, 2016. CBEST. Version 2.0.
3. NCSC (National Cyber Security Centre). Vulnerability management: Guidance to help organisations assess and prioritise vulnerabilities. `https://www.ncsc.gov.uk/guidance/vulnerability-management.`, 2016. NCSC. (Accessed: 05-03-2020).
4. CERT-UK. An introduction to threat intelligence, 2015. CERT-UK. TLP White.
5. CREST. What is Cyber Threat Intelligence and how is it used?, 2019. CREST. CTIPS (CREST Threat Intelligence Professionals).
6. Maryam Farsi, Alireza Daneshkhah, Amin Hosseinian Far, Omid Chatrabgoun, and Reza Montasari. Crime data mining, threat analysis and prediction. In *Cyber Criminology*, pages 183–202. Springer, 2018.
7. Forcepoint. What is Threat Intelligence?: Threat Intelligence Defined and Explored. `https://www.forcepoint.com/cyber-edu/threat-intelligence.`, 2020. Forcepoint. (Accessed: 29-02-2020).
8. Recorded Future. What Is Threat Intelligence? `https://www.recordedfuture.com/threat-intelligence/.`, 2020. Crowd Strike. (Accessed: 17-02-2020).
9. Intel & Analysis Working Group. What is Cyber Threat Intelligence?. `https://www.cisecurity.org/blog/what-is-cyber-threat-intelligence/.`, 2020. CIS (Centre for Internet Security). (Accessed: 26-01-2020).
10. Gerard Johansen. *Digital forensics and incident response: an intelligent way to respond to attacks*. Packt Publishing, 2017.
11. B Kime. Threat intelligence: Planning and direction., 2016. SANS Institute. White Paper.
12. Rob McMillan. Definition: threat intelligence. *Retrieved March*, 29:2019, 2013.
13. Reza Montasari. An ad hoc detailed review of digital forensic investigation process models. *International Journal of Electronic Security and Digital Forensics*, 8(3):205–223, 2016.
14. Reza Montasari. Formal two stage triage process model (ftstpm) for digital forensic practice. *Int. J. Comput. Sci. Secur*, 10:69–87, 2016.
15. Reza Montasari. Review and assessment of the existing digital forensic investigation process models. *International Journal of Computer Applications*, 147(7):41–49, 2016.
16. Reza Montasari. An overview of cloud forensics strategy: Capabilities, challenges, and opportunities. In *Strategic Engineering for Cloud Computing and Big Data Analytics*, pages 189–205. Springer, 2017.
17. Reza Montasari. A standardised data acquisition process model for digital forensic investigations. *International Journal of Information and Computer Security*, 9(3):229–249, 2017.
18. Reza Montasari. Testing the comprehensive digital forensic investigation process model (the cdfipm). In *Technology for Smart Futures*, pages 303–327. Springer, 2018.
19. Reza Montasari and Richard Hill. Next-generation digital forensics: Challenges and future paradigms. In *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, pages 205–212. IEEE, 2019.
20. Reza Montasari, Richard Hill, Victoria Carpenter, and Farshad Montaseri. Digital forensic investigation of social media, acquisition and analysis of digital evidence. *International Journal of Strategic Engineering (IJoSE)*, 2(1):52–60, 2019.
21. Reza Montasari, Richard Hill, Farshad Montaseri, Hamid Jahankhani, and Amin Hosseinian-Far. Internet of things devices: Digital forensic process and data reduction. *International Journal of Electronic Security and Digital Forensics*, 2019.

22. Reza Montasari, Richard Hill, Simon Parkinson, Pekka Peltola, Amin Hosseinian-Far, and Alireza Daneshkhah. Digital forensics: Challenges and opportunities for future studies. *International Journal of Organizational and Collective Intelligence (IJOCI)*, 10(2):37–53, 2020.
23. Reza Montasari, Amin Hosseinian-Far, and Richard Hill. Policies, innovative self-adaptive techniques and understanding psychology of cybersecurity to counter adversarial attacks in network and cyber environments. In *Cyber Criminology*, pages 71–93. Springer, 2018.
24. Reza Montasari, Amin Hosseinian-Far, Richard Hill, Farshad Montaseri, Mak Sharma, and Shahid Shabbir. Are timing-based side-channel attacks feasible in shared, modern computing hardware? *International Journal of Organizational and Collective Intelligence (IJOCI)*, 8(2):32–59, 2018.
25. Reza Montasari and Pekka Peltola. Computer forensic analysis of private browsing modes. In *International Conference on Global Security, Safety, and Sustainability*, pages 96–109. Springer, 2015.
26. Reza Montasari, Pekka Peltola, and David Evans. Integrated computer forensics investigation process model (icfipm) for computer crime investigations. In *International Conference on Global Security, Safety, and Sustainability*, pages 83–95. Springer, 2015.
27. NormShield. Cyber Threat Intelligence. `https://www.normshield.com/cyber-security-with-artificial-intelligence-in-10-question/.`, 2020. Recorded Future. (Accessed: 24-02-2020).
28. John Pescatore. SANS Top New Attacks and Threat Report, 2019. SANS Institute Cyber Security Report.
29. Z. Pokorny. 4 Ways Machine Learning Produces Actionable Threat Intelligence. `https://www.recordedfuture.com/machine-learning-threat-intelligence/.`, 2018. NormShield. (Accessed: 25-01-2020).
30. Dave Shackleford. Who's using cyberthreat intelligence and how? *SANS Institute. Retrieved January*, 24:2018, 2015.
31. Dunlap Stephen, Riceand Mason, Mills Robert, and Sibiga Matthew. Applying cyber threat intelligence to industrial control systems. *Journal of Cyber Security and Information Systems*, 7(2), 2016.
32. Crowd Strike. Cyber Threat Intelligence. `https://www.crowdstrike.com/epp-101/threat-intelligence/.`, 2019. Crowd Strike. (Accessed: 27-02-2020).