CTI, CTI, CTI: Applying better terminology to threat intelligence objects

Adam Greer

CTI, CTI, CTI: Applying better terminology to threat intelligence objects

GIAC (GSEC) Gold Certification

Author: Adam Greer adam.greer@sans.edu
Advisor: *Clay Risenhoover*

Accepted: December 13, 2020

Abstract

Increased awareness of the need for actionable cyber-threat intelligence (CTI) has created a boom in marketing that has flooded industry publications, news, blogs, and marketing material with the singular term applied to an increasingly diverse set of technologies and practices. In 2015, Dave Shackleford and Stephen Northcutt published findings of a survey sponsored by some of the largest names in cyber-threat intelligence at the time in order to address the widespread confusion around what precisely cyber-threat intelligence is and how it is generated, delivered, and consumed. In this research, they note that "...a shortage of standards and interoperability around feeds, context, and detection may become more problematic as more organizations add more sources of CTI..." (Shackleford, 2015). However, IT security teams have matured drastically since then, and most research has been applied to automation and standards for specific sub-domains, such as dissemination. This paper analyzes the current CTI environment and uses a defined methodology to develop a taxonomy for the domain that clarifies the application of CTI to security programs and serves as a foundation to further domain research.

1. Introduction

Intelligence gathering is a crucial factor in the security management process. Aggregating data for analysis is one of the steps in the intelligence cycle which consists of planning, collecting, processing, analysis, and dissemination of data. The cycle is designed to identify sources of raw data that can be analyzed into intelligence, or "information relevant to decision-making" (FBI, 2020). The application of the intelligence cycle with the intent to produce actionable threat intelligence is a well-defined process. However, the unique industry jargon and level of technical acumen required to effectively apply the threat intelligence lifecycle to cyber systems separates the processes sufficiently enough to require the distinction of cyber-threat intelligence processes.

The limitation of resources demands efficient methods of gathering data for analysis as part of the intelligence development process. This demand has driven the development of an industry dedicated to providing useful raw and processed data to meet the needs of organizational cyber-threat intelligence requirements. However, the classification of objects in the cyber-threat intelligence domain into a taxonomy has not been previously addressed, and these services and systems were introduced using largely ad-hoc nomenclature (Nickerson, 2013).

Taxonomies play an essential role in research and management because the classification of objects helps researchers and practitioners understand and analyze complex domains. This universal characteristic of taxonomies is also highlighted by Miller & Roth (1994), who noted that "taxonomies … are useful in discussion, research and pedagogy." McKnight & Chervany (2001) state that taxonomies can order otherwise disorderly concepts and allow researchers to postulate on the relationships among the concepts. The purpose of this research is to develop and present a taxonomy that can be used to clarify distinct groups of CTI objects and therefore provide a defined nomenclature that can be used to facilitate meaningful conversations regarding the selection and usage of CTI objects in support of an organization's security program. The formal nomenclature will also provide a scaffold for further analysis of the design

principles and contextual relationships between the artifacts of the domain which can further the understanding of the industry's successes and shortcomings.

2. Environment Exploration

The Information Age has resulted in complex infrastructures that are designed with the sole purpose of delivering data in massive quantities. Cyber-threat intelligence was not immune to this unprecedented surge in readily available data, and the general market place was deluged in a diverse set of streams of raw and analyzed data that required specialists with specific skill sets to properly interpret and analyze in order to become useful intelligence within the context of their circumstances.

A search for the term on common engines returns results from the top names in the cybersecurity industry. Results include sites dedicated to explaining what cyber-threat intelligence is, why it's crucial, quick lists of general types of indicators of compromise or attack. The explanations are followed by a discussion of how that vendor conducts their research and can save time and money by utilizing the service or participating in the sharing network. Each result has a unique degree of depth and focus, depending on what the end goal is. CrowdStrike focuses on advanced persistent threats (APTs) and thoroughly explains the threat intelligence lifecycle geared toward dedicated threat intelligence teams tracking nation-state threats. Cisco focuses on leveraging machine learning capabilities in analysis processes to compensate for limited human analyst resources. The Center for Internet Security focuses on the importance of sharing threat intelligence through its Multi-State Information Sharing and Analysis Center. Industry blogs and media sites provide lists of "The Top Cyber Threat Intelligence Feeds" and reviews of various solutions. A comprehensive EC-Council explanation ends with an offer for obtaining the EC-Council Certified Threat Intelligence Analyst certification.

Scholarly search engines, such as Google Scholar and EBSCOhost, provide articles and papers on innovative techniques and tools used for maturing CTI processes. The methods to incentivize threat intelligence sharing using blockchain, analysis of neural networks and the operational processes best geared for artificial intelligence methods, and tools for automatically extracting and analyzing a specific CTI feed with

artificial intelligence or machine learning comprise the majority of results. These results reference the CTI domain as the umbrella under which their unique research is being conducted, but no further reference is made. The remainder of the returned research is primarily analysis of the industry wide usage, sharing, privacy concerns, and application of CTI. The commonality between the two types of scholarly results is the papers are laden with technical jargon and expectations of prerequisite knowledge in several fields of computer science, thereby significantly limiting their accessibility.

The problem is further compounded by freshly discovered vulnerabilities and the constant drive by vendors to market and coin buzzwords, simple phrases that represent incredibly complex ideas or difficult tasks. The influx of marketing terminology in combination with the rapid availability of information has caused the CTI consumer base to be introduced to CTI that is not directly actionable on their part. An example of this is the disclosure of the Spectre and Meltdown vulnerabilities. The specification of a unique moniker made the intelligence easy to consume for personnel that lacked the knowledge and experience to contextualize the information and make a proper risk calculation. The subsequent media surrounding the vulnerabilities seemed to overlook the small detail that in order to exploit the numerous weak points in the microarchitectural decision-making process, you had to be root on the local system (Kocher, 2018).

This prerequisite significantly limited the overall risk of the vulnerability being successfully exploited by an unauthorized malicious actor; after all, most cyber defense mechanisms are aimed at preventing unauthorized users from gaining that level of access to a system. However, that caveat was lost in the noise and the response from the general media, and therefore the public, was that the world was going to end with the whole Internet being hacked and critical public infrastructure shut down and held ransom. By applying monikers to the vulnerabilities, the researchers increased the availability of the intelligence to a much wider consumer base than was capable of accurately assessing its impact. Headlines such as ‘Spectre and Meltdown: hackers can steal all your data’ have a much wider audience than headlines such as ‘New exploit chain for predictive execution branch manipulation through the use of transient instruction customization was

discovered and disclosed by leading security researchers' and therefore panic was induced for a wider percentage of the population than was necessary.

This permeation of buzzwords and over-simplified references leads to organizations "spending money on data that they do not need or are unable to utilize" and "subscribing to threat feeds that do more harm than good" (Lee, 2020). This note is negligibly different from the SANS 2014 Analysis and Intelligence report that resulted in Dave Shackleford (2015) commenting, "There's a lot of confusion around what threat intelligence is and how it's delivered and consumed.". Despite the advancements made in the collection, processing, and dissemination of methods and technologies and increased awareness of CTI's importance to an information security program, the terminology used to isolate and identify useful CTI sources is lacking.

The $(ISC)^2$ 2019 Cybersecurity Workforce Study determined that the majority of organizations participating in the survey, 65%, reported a shortage of internal cybersecurity staff. The Cybersecurity Workforce Study is a survey of organizations that currently employ cybersecurity staff; therefore, it can be extrapolated that since organizations that have identified a need for cybersecurity personnel cannot fill all identified required positions there are many organizations with no cybersecurity specific staff at all. Likewise, if the organizations that possess Information Security departments mature enough to have dedicated CTI teams are unable to quantify the value of the CTI they collect, it can be deduced that the less mature organizations without dedicated specialists to understand what CTI products are actually delivering have little chance of accurately identifying and selecting the CTI sources that are best suited for their organization.

Gartner (2018) predicts that the global information security market is poised to reach $170 billion in 2022, and Statista (Holst, 2020) estimates $248 billion by 2023. The latest report by Cybersecurity Ventures (Morgan, 2020) estimates that the global cyber-crime market will top $6 trillion by 2021. In a general context, these numbers imply that information security teams will have 24% of the budget that the cybercriminals will have. To state this problem simply, the average network defenders are short-handed and

outgunned, and the market is flooded with ambiguous literature, adding confusion to the areas that need the most clarity to most effectively utilize limited resources.

2.1. CTI Industry literature review

To identify commonly accepted nomenclature currently in use by the CTI industry and the surrounding verticals invested in the industry the author chose to perform an multivocal literature assessment. These types of reviews are generally not performed in lieu of more academic-focused analysis due to the inclusion of data that is experience and opinion based, commonly referred to as grey literature (Garousi, 2016). These publications are colloquially called blogs, marketing, and general media. However, given the existing nomenclature developed to support existing domain objects are either innate to the threat intelligence lifecycle or unique to specific type of intelligence being operated on or the sponsor of the object, the grey literature was likely to provide a more accurate depiction of accepted and established terminology used by the average consumer. The inability to rely on strictly academic and industry leader publications is a symptom of the development of largely ad-hoc development of nomenclature in the information systems field (Nickerson, 2013).

This review identified existing terminology such as ‘source’ and ‘report’ and other formal terms innate to the concept of intelligence and that mirror classical threat intelligence terminology. Subjective analysis of nuances differentiating terms like 'feed' and 'source’ resulted in the identification of characteristics. More active terminologies in use, such as collection networks or vulnerability disclosure feeds, are too broad to be used once applied to the selected dimensions. However, the equally vague term of 'aggregate vulnerability databases’, the National Vulnerability Database for example, appears to fit within the same dimensions. This anomaly is explained by the breadth of particular characteristics that are unique within the context of the applied dimension. The nuances of each characteristic within each domain will be elaborated on in Section 3 as certain familiar terms will have specific constraints applied to their interpretation resulting in a unique contextual definition.

Academic research papers and industry publications regarding the CTI domain are reflective of Nickerson’s (2013) and Mavroeidis’ (2017) findings of the field being

largely ad-hoc development. For example, the Common Vulnerability Scoring System (CVSS), managed by FIRST (2020), provides a framework for communicating the characteristics and severity of vulnerabilities. It uses terms such as Base, Temporal, and Environmental as variable descriptors for determining the score to compare a vulnerability's potential impact against other vulnerabilities. The Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) Framework (Alberts, 1999) also focuses on determining the impact of vulnerabilities on assets, but nowhere in OCTAVE's documentation is there a reference to Base, Temporal, and Environmental labels for processing vulnerability data to determine impact. For this reason, very little unified terminology was able to be derived from this group of literature.

Vendor publications and research are mostly focused on the CTI industry rather than the CTI domain. Aimed at decision makers primarily focused on budgetary ramifications, these documents use superfluous marketing language rather than hard technical jargon. Primarily consisting of problem-solving solutions and landscape-level reports, this literature contains language narrowly applicable to a product. When viewed as an aggregate pool of resource documents the Application Area dimension was identified.

Since the information age has introduced a multitude of cyber systems into the common routine, even the general public consumes cyber threat intelligence at a broad level. This literature is largely categorized into Media and Entertainment publications and productions. These sources provide a narrow but comprehensively catastrophic outlook on what could happen should personal or national security be breached. This literature tends to use terminology and nomenclature from the technology field bordering on the common body of knowledge, such as password and email, and therefore provided little extra in the way of providing nomenclature specific to the domain.

3. Methodology for Taxonomy Development

Users, researchers, and developers need to be able to know where a new object fits within a domain in order to determine if it is something entirely new and unique, a significant variation of an existing application, or just a retread of what already exists. As

Iivari (2007) explains, “The research goal at the conceptual level is essentialist: concepts and conceptual frameworks at this level aim at identifying essences in the research territory and their relationships.” A taxonomy provides a basis for making this determination and highlights voids where new avenues of research might be developed. This paper is focused on using the characteristics of the objects being examined to identify groups of objects, and therefore will use a phenetic methodology. The phenetic approach applied to the CTI domain highlights the characteristics of the sources of that data or intelligence into the appropriate section of the intelligence life cycle as it functions as a decision support system for the organization (Nickerson, 2013).

A taxonomy represents distinct objects comprised of mutually exclusive and collectively exhaustive characteristics so each object under consideration has a single characteristic for each dimension (Nickerson 2013). The mutually exclusive restriction means that no item can have two different characteristics in a dimension. The collectively exhaustive restriction means that each object must have one of the characteristics in a dimension. Together, these conditions mean that each object has exactly one of the characteristics in each dimension. These constraints are implemented in order to develop a useful taxonomy as part of a search process but also require the taxonomist to create nuanced nomenclature. The resulting matrix formation of listing objects and marking which characteristics are displayed from a dimension allows the foundational objects to adapt as the domain changes. The design lays the foundation for the ability to infer properties of an object based upon membership in a particular class, which should prove useful to researchers conducting further domain analysis.

The development process requires identifying the audience of the taxonomy and specifying a projected use for the taxonomy. This involves the taxonomist theorizing who the users could be and deciding, based on experience and empirical evidence, what the users could do with the taxonomy. The audience for this taxonomy is intended to be information security specialists and academic researchers, but in order to fulfill the primary purpose of a taxonomy, a conscious effort was made to develop a nomenclature that would be consumable by information technology professionals and managers. This choice of nomenclature was utilized in order to support the projected use of the

taxonomy, which is to supply nomenclature for use in providing clarity around what CTI is and how it should be applied to an organization's information security program.

The methodology defined by Nickerson also includes specific ending conditions that test the taxonomy during development. This approach is consistent with the design science 'generate/test cycle' (Nickerson, 2013). There are both objective and subjective ending conditions which are designed to provide assurances that a comprehensive set of data was analyzed in order to produce distinct and useful nomenclature. The primary ending condition is that the taxonomy must satisfy the definition of a taxonomy; in that, it has mutually exclusive dimensions and collectively exhaustive characteristics. Nickerson (2013) identifies eight additional objective ending conditions that will be used in this research. The conditions listed below were adapted from Sowa & Zachman's (1992) rules for Information System architecture frameworks:

- All objects or a representative sample of objects have been examined.
- No object was merged with a similar object or split into multiple objects in the last iteration.
- At least one object is classified under every characteristic of every dimension.
- No new dimensions or characteristics were added in the last iteration.
- No dimensions or characteristics were merged or split in the last iteration.
- Every dimension is unique and not repeated (i.e., there is no dimension duplication).
- Every characteristic is unique within its dimension (i.e., there is no characteristic duplication within a dimension).
- Each cell (a combination of characteristics) is unique and is not repeated (i.e., there is no cell duplication).

Subjective ending conditions are necessary conditions for a useful taxonomy and are essentially iterative checks to ensure that the taxonomy is concise, robust, comprehensive, extendible, and explanatory. These conditions are the minimal subjective requirements that must be met for the method to terminate. These conditions result in

new dimensions being added to the process, a restart of the iterative process, characteristic identification, or the refinement of nomenclature. For example, during the iterative process, an initially identified object, Detection alerts, was later refined into Indicators of attack (IOA) detections and indicators of compromise (IOC) detections due to the addition of the Application Area dimension. The application of different ending conditions applied to the same data will generate different taxonomies as the subjective opinions result mostly from the taxonomist searching for useful, not necessarily optimal, solutions for their primary objective.

With the audience, use, and termination conditions defined, the taxonomist can initiate functional processing with either an empirical approach or a conceptual approach (Nickerson 2013). Determining the appropriate approach to use depends on the availability of data about objects under the domain and the taxonomist's knowledge of the domain of interest. The empirical-to-conceptual approach requires the taxonomist to identify a subset of objects that they wish to classify from an environment with pre-existing published documentation regarding the domain. The opposite conceptual-to-empirical approach requires the taxonomist to have an intimate understanding of the domain, but little information regarding the domain has been published. There has been an established CTI industry for four decades, so the conceptual-to-empirical approach is not applicable in this context. Accordingly, this paper will use the empirical-to-conceptual approach to identify characteristics or pre-existing CTI objects. Since the CTI domain is well established with producers, consumers, vendors, clients, and dedicated academic efforts, the amount of distinct domain-related objects from which to draw terminology is extremely susceptible to what Aldenderfer & Blashfield (1984) refer to as "naive empiricism." This occurs when an inexperienced taxonomist attempts to analyze large disparate terminology in the hopes that patterns will emerge. While functionally processing object sets through the methodology, it became clear that avoiding this naivety would be difficult.

The choice of the characteristics in a taxonomy is a central problem in taxonomy development (Nickerson, 2013). An overarching meta-characteristic must be specified at the beginning of the development process. This meta-characteristic is the most

comprehensive characteristic and serves as the basis for the selection of objects in the taxonomy and establishes the direction for the entire analysis of artifacts within the domain. The meta-characteristic for this taxonomy was chosen to produce nomenclature such that it would be immediately useful in application and potentially shine a light on previously overlooked sources of critical threat intelligence. For this foundational taxonomy, all artifacts analyzed were described as CTI in material referencing, describing, or documentation regarding the functionality of the artifact.

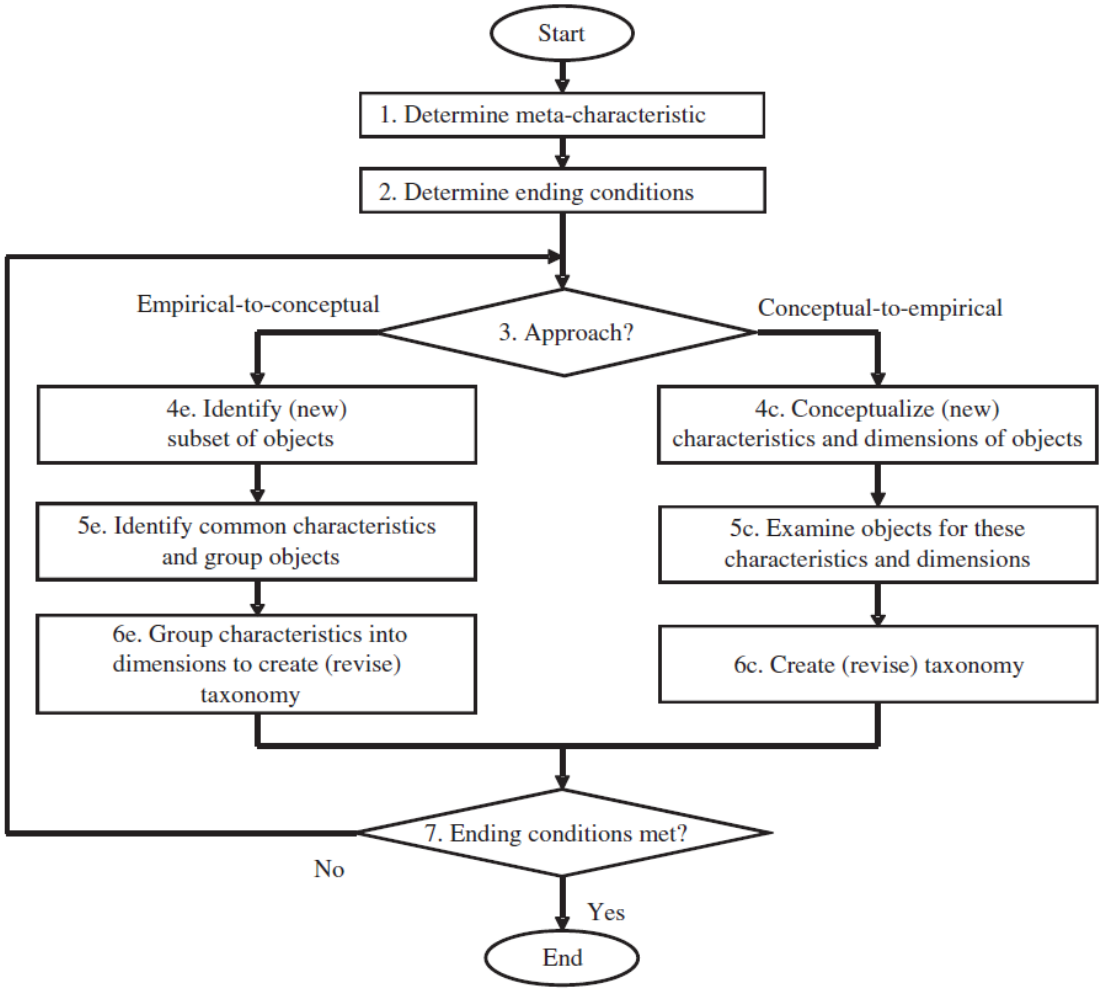

Figure 1 The taxonomy development method. (Nickerson, 2013)

3.1. Methodology application to CTI domain

As stated before, the classification and grouping of objects in a domain of interest into a taxonomy is a fundamental problem in many disciplines. Developing a taxonomy is a complex process that has not been adequately addressed in the information systems

literature on information systems, and taxonomy development in the information system field has largely been ad hoc (Nickerson, 2013). This has led to a limited pool of resources related to the formal grouping and definition of CTI objects as they pertain to the domain, but a diverse pool of resources relating to an informal defining of CTI objects.

3.1.1. Dimension – Audience

The functional processing of CTI objects through the empirical-to-conceptual methodology described above was started by taking objects identified through searches and using the experiential knowledge gained through the author's experiences in the information security field to identify who would be the best recipient for a particular object or the target audience.

The target audience in this context refers to the grouping of roles where the intelligence would have the most significant impact. For instance, malicious IP lists are referenced by SOC analysts and specialists on a daily basis, but that same list has the most significant impact when given to the administrator who can implement that list on a device with IP filtering capabilities. For the sake of simplicity, the author chose to base the audience characteristics on universal IT Support Tier role definitions since they are well known. This dimension could be further divided into more specific roles, but this level of division did not seem to serve a useful function at this juncture of the taxonomy development process.

The first and broadest characteristic group is administrators. These consumers are the ones responsible for implementing system configuration changes. These are roles that are more closely associated with IT than cybersecurity. Common roles and titles associated with this group are technician, engineer, administrator, and Help Desk manager. Software developers are also included in this group as there is no significant difference between a firewall administrator updating a configuration, a server administrator pushing updates to clients, and a developer changing a program to stop using a deprecated function.

The second group is SOC analysts. These consumers are the ones responsible for operating and maintaining internal CTI collection and alerting processes and systems. These are individuals who are primarily responsible for applying external intelligence to the unique circumstances of the organization and monitoring for anomalies in internally generated intelligence feeds. Generally the first line of defense of an organization, these personnel are responsible for reviewing all the CTI an organization ingests and generates.

The third group is specialists. These consumers are Level 3 and 4 Information Security specific roles such as, but not limited to: forensic investigators, penetration testers, security consultants (client-facing engineers and managed security service providers), and dedicated CTI professionals. These individuals are intimate with information security theory and practices, and are responsible for protecting large enterprises, multiple organizations, nations, and even the Internet as a whole.

The fourth group is the directors. These consumers direct system changes and have responsibilities that fall between the upper echelon of management to the executive level. These individuals are responsible for advising Executive C-suites or Boards of Directors on corporate strategy and budget considerations as they relate to the enterprise risk management program. This can effectively be considered the personnel who have direct responsibility for ISMS program management.

3.1.2. Dimension – Application Area

This dimension was highlighted by the characteristics of the area that the intelligence is best applied to in a traditional Vulnerability Management program. This dimension will help clarify some of the confusion around where CTI comes from and how it should be utilized. In the SANS 2020 CTI report (Lee, 2020), only 5.0% of the respondents believed that they produced raw threat data, and 6.6% produced contextual threat alerts. These responses realistically mean that respondents were not commercially producing CTI for others to consume, but it could imply that organizations are too focused on ingesting data from external sources and are missing critical information produced from their own assets.

The first and second characteristics of this dimension are Hardware and Software. These terms are simple to recognize and apply. Hardware refers to any electronic device that runs specialized firmware developed by a vendor. Hardware includes, but is not limited to: switches, routers, firewalls, wireless access points, cable modems, televisions, IoT/ICS/SCADA devices, processors, motherboards, GPU cards, cameras, RFID access systems, gate control systems, vehicles, medical devices, printers, and game consoles. Software refers to any application that runs directly on top of a piece of hardware or another piece of software. Examples include operating systems, local applications, web applications, scripts, and the programming languages used to create them.

The third characteristic, Network, requires an explanation of the particular confines in which the term operates. This singular term identifies the connection between systems in which vulnerabilities exist in the act of participation rather than a specific hardware or software component. For example, the vulnerabilities innate to the TCP/IP stack don't come under consideration until a connection is established between two systems. The risk of a message being intercepted is not a vulnerability that can be mitigated directly. The impact of it being intercepted can be mitigated, but it cannot be 100% prevented as it traverses the Internet. Likewise, participating in the network introduces its own vulnerabilities such as denial of service.

The fourth characteristic group is Organizational Security. This characteristic is qualified by intelligence that is uniquely applicable to an organization's risk management decisions. Help Desk ticket trends, aggregator platform dashboards, control frameworks, threat simulation reports (penetration tests, tabletop exercise after-action reports, red team exercises) are several CTI objects that exhibit this characteristic. These CTI objects source data for consumption by personnel attuned to an environment's idiosyncrasies. What would be a catastrophic compliance report for one organization could be considered a glowing pass for another. A number of requests to the Help Desk asking for help with rejected email delivery might indicate an email compromise for one organization but be a normal traffic pattern for another.

The fifth and final characteristic is Landscape. This refers to the intelligence that has been processed from large data sets that cover multiple other spectrums tracking the

activity of advanced persistent threats (APTs). Objects that exhibit this characteristic include aggregate intelligence databases, like cvedetails.com, that collect and organize data for consumption on demand. Threat actor analysis reports that track the details of how specific groups are conducting cyber-warfare. Landscape analysis reports generally involve an analysis of survey results or extra-domain data through an information security lens. Examples of the landscape analysis reports are the familiar annual industry and vendor reports that share their analysis of metrics collected from customers participating in sharing programs such as the (ISC)2 Workforce Study, the SANS Security Awareness Report, the Verizon DBIR, Gartner's Magic Quadrants, Crowdstrike's Cybersecurity Report, ISACA's State of Cybersecurity, and Cisco's Cybersecurity Report. It may also include reports on the industry itself such as this paper which is applying knowledge and methodologies from the linguistic field of study to the information security domain.

3.1.3. Dimension – Functional Level

The third dimension formulated was the managerial decision level to which the intelligence is best suited. The characteristics of this dimension are tactical, operational, and strategic. This dimension is supplementary to the first dimension as a necessary distinction to which process level the intelligence should be applied to.

Operational processes are those that require constant maintenance and are mostly automated in the current environment. Security companies have honeypot networks that capture raw attack traffic and automatically convert it into a detection for an IDS or SIEM platforms. Organizations rely on detections and patches being automatically downloaded and installed. In traditional threat intelligence, this is a much more complex process that would likely involve field agents embedding themselves in enemy territory and reporting data back to HQ or scouts radioing updates about enemy movements. In cyber-threat intelligence collection, these non-automated tactical processes would involve tasks like resolving support tickets, responding to an incident, reviewing aggregator platform dashboards, and other similar tasks.

Tactical processes are those processes that keep the organization functioning and growing by executing the output of the Strategic level. Managerial decisions regarding

the workforce, business development, and growth projects, reviewing support ticket trends for signs of larger issues, determining acceptable levels of system baseline hardening and framework implementation, and researching vulnerabilities to determine contextual risk ratings are all examples of tactical processes.

Strategic processes are those processes that guide the industry and an organization's overall security strategy. Within the context of the domain, these processes determine the level of organizational risk based on landscape reports, ensuring that appropriate intelligence requirements for internal collection systems are clearly identified and are changed as the business requirements change. It also ensures that resources are available to effectively fulfill the collection and analysis process of those requirements, and ensuring that the environment has sufficient knowledge and capabilities to accurately defend against identified threats.

3.1.4. Dimension – Implementation

This dimension is not immediately apparent due to cyber-threat intelligence's direct association with technology and society's increased reliance and interaction with it. However, it might be one of the more useful dimensions in this taxonomy as it clearly identifies what type of analyst, computer, or human, should be receiving the intelligence object to process. This dimension became apparent due to the distinct difference between sources providing 'feeds' and 'reports.'

Intelligence processing can be more formally described as a decision support system. Data is gathered and analyzed in order for leadership to make better decisions. As theorized by Alter (1977), decision support systems generally fall into seven categories. Alter approached this topic through an early business computing lens of primarily accounting data processing, so not all of Alter's taxonomy is applicable to the CTI domain. The modern era has implemented advanced algorithms to perform some of these decisions without human interaction. This has led to a separation of intelligence implementations of operational-level electronic data processing (EDPs) and tactical and higher-level decision support systems (DSSs) that require human contextualization. Lee (2020) refers to these as "processing tools" and "management tools."

3.1.5. Dimension – Direction

There are two characteristics of this dimension: uni-directional and bi-directional. Uni-directional objects are synthesized by external entities and are delivered through sharing mechanisms that are non-tunable by the consumer and are meant to be contextualized by the consumer. Bi-directional objects are either synthesized using data sourced from non-public shared datasets in which only the contributors to the dataset receive the intelligence or intelligence that is utilized to modify collection mechanisms and requirements. Uni-directional objects are the familiar cyber-threat intelligence transactions such as AV signature updates, bad IP updates, Indicators of Attack detection feeds, hardened baselines, and aggregate vulnerability databases. Bi-directional object examples consist of support ticket trends, IOA and IOC feeds, aggregator platforms, industry-specific threat sharing networks, and landscape reports.

3.2. Proposing the CTI Taxonomy

The following nomenclature was derived by processing existing CTI domain objects through the methodology as described above. Much of this terminology will be intuitively understood by the intended audience; however, each term will be accompanied by a definition and known examples of that object. A tabular format of the taxonomy highlighting the object characteristics is attached as Appendix A.

Aggregate vulnerability databases are repository systems that uniquely identify and record details of all known vulnerabilities. These databases serve as reference points for consumers who are investigating risk profiles of systems that are relevant to a particular task. Well-known aggregate vulnerability databases are the National Vulnerability Database found at nvd.nist.gov and CVEdetails.com.

Aggregator Platforms are applications that analyze multiple sources and produce alerts based on input. These platforms generally include dashboards and graphs to assist in the analysis process. Examples are common SIEMS such as SolarWinds, Splunk, Leargas, LogRhythm, and vendor-specific "security fabrics."

Anti-virus Detection Alerts are alerts triggered by anti-virus or anti-malware software. These applications monitor endpoint systems for anomalous activity that could be the result of malicious unauthorized activity.

Compliance Audit Report objects are bi-directional objects that are sourced from constant reviews of the landscape but applied to organizational entities. These reports facilitate an understanding of an environment's adoption of the control actions identified by the cybersecurity industry as critical in maintaining an information security system.

Control Frameworks are sourced from expert analysis of appropriate defensive mechanisms that should be applied to protect electronic systems. Control Frameworks exists for multiple sets of data types such as HIPAA for medical data, PCI-DSS for credit card data, the CIS CSC for general data protection, and NIST 800-53 for federal data. There are many other control frameworks already defined in the CTI environment.

DNS allow/block-lists are the aggregated data from monitoring systems and report fully qualified domain names that are associated with malicious activity or are known, trusted sites. Many of these lists are maintained and sourced by network appliance vendors, such as Cisco Umbrella and FortiNet's Botnet C&C DNS filter, but there are many open-source lists such as the Malware Domain List and the SpamHaus Project.

DNS Information Platforms are a form of aggregate databases that record WHOIS information and other data specific to a fully qualified domain name's (FQDN) history. RISKIQ's platform is an example of this type of object.

Exploit PoC databases are specialized intelligence databases that record proof of concepts for exploiting identified vulnerabilities in hardware or software. Originally this object type was classified as a feed due to many news media outlets that track and write articles about the latest exploits discovered. However, after some analysis, these objects proved to be selective in nature and therefore were not qualified to be classified as members of this group.

Adam Greer, adam.greer@sans.edu

Hardened Baselines are reports from security specialists that review common systems and document recommendations for limiting the attack surface of a system. Most operating systems and applications are configured by default with options that are not required for common users, thereby increasing the attack surface. Examples of this object include the CIS benchmarks, Windows security baselines, and custom-designed system images created for a specific environment.

Hardware End-of-Life objects can be delivered from aggregate databases, such as EOL tracker at eoltracker.com, but are much more effective when sourced from the manufacturer ahead of time for use in strategic planning efforts for lifecycle replacement schedule development.

IOA Detection objects alert analysts to attacks against infrastructure. These indicators reflect known attack traffic patterns and can be sourced from external collection networks using honeypot data, packets logged on a firewall policy that should not have any matching packets, or vendor services including anti-virus detections.

IOC Detections objects alert analysts to systems exhibiting signs of compromise. These are specific items that would indicate compromise to a forensic analyst or incident responder and consist of registry entries made by known malware, socket data used for command and control, and logs triggered by unique malicious TTPs.

IP allow/block-lists are similar to DNS allow/block-lists but pertain to IP addresses rather than FQDNs.

IP information Platforms are similar to DNS information platforms and can provide similar data, but the sorting key for the data is the quad octet identifier and not the FQDN. There is also unique data found in these platforms, such as port information.

Landscape Analysis Reports take extremely diverse sets of data gathered from sources representative of multiple industries, regions, sizes, and organization type. The resulting synthesized intelligence has global implications. These reports include industry solution analysis, such as Gartner ratings, to control framework updates, such as NIST 800-53

revisions 2-5, and other reports such as the Verizon and Secret Service's DBIR and SANS Security Awareness report.

Malware Identification Platforms are objects that perform analysis against items and return analysis of the item's actions and characteristics as compared to known malicious behaviors. These platforms are supplementary to local malware identification, as they are not installed locally and only provide analysis on demand. HybridAnalysis and CrowdStrike's Falcon Sandbox are examples of this object.

Network Traffic alerts are objects from sensors dedicated to reviewing traffic anomalies and metrics. These objects are commonly fed to aggregator platforms for visual contextualization or ticketing platforms meant to alert appropriate personnel. Sudden changes in bandwidth consumption, traffic on unexpected ports, endpoint to endpoint traffic, or malformed protocol headers are all examples of this object.

Software end-of-life objects can be delivered from aggregate databases, such as the SANS EoL software list, but are much more effective when sourced from the developer ahead of time for use in strategic planning efforts for acquisition budgets.

Support Ticket Trend objects are uniquely sourced and consumed objects that only synthesize intelligence applicable to a singular environment. These objects are extremely useful for identifying widespread availability failures and other events that indicate compromises of security controls and mechanisms.

Threat Actor Feeds are highly specialized collection mechanisms that track the actions of a singular identified global threat actor in real-time. These objects include collection mechanisms employed by organizations that can actively track the global activity and attribute it to specific APT groups: Mandiant, Varonis, FireEye, and the MITRE Corporation.

Threat Actor Reports are the published analysis of the data collected through Threat Actor Feed mechanisms. These reports enumerate current targets, tactics, techniques, and procedures of specific threat actor groups currently active in global electronic warfare.

Adam Greer, adam.greer@sans.edu

Threat Simulation Reports objects are currently referred to as "security assessments" or "penetration test reports." The objects provide intelligence to decision-makers regarding the current security posture of an organization based on the effectiveness of their current controls in stopping known attack methods, as well as a demonstration of an impact when those controls fail.

Update/Hotpatch Feeds are objects that are designed to close known vulnerabilities as soon as possible while maintaining functionality. These objects are almost exclusively provided by manufacturers and developers but require interaction on behalf of Administrators to ensure that systems are set to utilize these feeds and that they are successfully implemented.

Vulnerability Scan Reports are objects that provide organization-specific results and provide critical context around aggregate vulnerability databases, end of life trackers, and update/hotpatch feeds that are immediately actionable to Administrators. These objects are sourced from vulnerability scanners such as Qualys, Nessus, and OpenVas.

4. Discussion

4.1. Experiential application of the proposed taxonomy

The meta-characteristic used to contextualize the existing objects within the domain was chosen to fulfill the primary objective of a taxonomy, which is to be useful. Therefore, the first theoretical application for the taxonomy is to be used in cyber-threat intelligence programs already functioning in organizations as a classification mechanism to identify what objects are effectively implemented and which are missing. The taxonomy will also provide clarity around what type of intelligence is being supplied or derived from a specific CTI object. This usage should have a direct application in most programs. As Lee (2020) noted, even organizations with dedicated CTI teams are experiencing confusion around what CTI is useful to their environment. This taxonomy should be applied to real-world environments and analysis performed of how useful the terms are and if any terms can be further refined.

4.2. Theoretical application of the proposed Taxonomy

The developed taxonomy could be applied in an industry-wide analysis of the number of producers of each object and where opportunities for streamlining the delivery of intelligence exist. This analysis using the proposed terminology as a focus could highlight the biases in the current environment towards developing robust production and ingestion systems instead of analysis and decision systems. This application could be applied to the entire CTI domain as a whole or for specific dimensions. The fourth dimension of this taxonomy that classifies objects as identified by EDPs and DSSs would be particularly subject to this analysis.

These terms could also be highlighted in surveys used to generate landscape reports to gain insights on how organizations are utilizing aggregator platforms and tracking internally generated threat metrics. Responses to these survey questions help reveal the source of the confusion in the industry as to what and how to apply CTI. These terms could also help clarify the disparate literature regarding the CTI industry.

4.3. Research Limitations

This taxonomy should be considered, at best, an analysis from a general practitioner of the cyberarts and not from the perspective of an expert linguist or CTI expert. The need for such a taxonomy became apparent to the author when the author's organization built a honeypot network that automatically takes the traffic it receives and converts it into a ZEEK detection. This is valuable data in many regards and can be used as intelligence for threat actor reports, organizational incident response qualifiers, and regional threat quantification metrics. The feed was given a proprietary moniker and a description of "Threat Intelligence Network." Therefore, this taxonomy should be reviewed and refined by specialists in the domain and those in the linguistic field with experience in taxonomy development.

An assumption was made early on in the process that there would be a clear nomenclature for processes used by security personnel at different levels of oversight that utilize different CTI objects. However, a dimension with unique characteristics was not

able to be identified during the iteration process. This is likely the result of the meta-characteristic used to drive the development process and the author's choice of Dimensions.

Another limitation of this proposal is an overwhelming dataset, a single researcher, and the time constraints placed on the development process. Despite being informed of the risk of "naive empiricism" (Alderberger & Blashfeld, 1984), the author succumbed to this practice in several early attempts due to the sheer number of sources of unique cyber-threat intelligence. The initial attempts at identifying characteristics of dimensions resulted in many hours surfing between pages indexed as "cyber-threat intelligence." While this practice was curtailed early on in the development process, it is possible the remnants of those iterations that were deemed salvageable had an unintended effect that limited the nomenclature selection process.

4.4. Implications for Future Research

As acknowledged previously, the initial dimension was left explicitly generalized for ease of consumption. Further refinement of this dimension could be made using the NIST NICE framework. This extremely specific breakdown of cybersecurity roles might introduce the most comprehensive nomenclature for unique objects. However, since such a comprehensive list would prove daunting for personnel attempting to establish a CTI program with a shortage of personnel, this research would be helpful to further academic research into the utilization of cyber-threat intelligence in the modern landscape.

Input from experts like GIAC Cyber-Threat intelligence professionals and others who are fully immersed within the domain is another crucial step in the development process. The author, while consistently immersed in the CTI domain in many aspects, is not acutely familiar with the nuances and expanses of the environment. Professionals in the CTI domain will provide a more experiential understanding and will be able to add objects missed by the author. Immediately notable examples that the author chose not to include due to a lack of nuanced understanding are IRC channels, black market sites and forums, and pecuniary tracking mechanisms used to identify cybercriminals or the inclusion of the general public into the audience dimension.

A researcher with a greater set of data could perform an analysis of specific objects to determine if the application of an analysis refinement dimension could be applied. This would classify objects by the level of analysis and refinement are performed on data sets prior to the delivery of the intelligence. A honeypot -based detection generator would be considered raw, while a report on the actions of a specific APT group would be quite refined. Were this dimension applied as the primary dimension it would likely generate a very different looking taxonomy.

5. Conclusion

The cyber-threat intelligence domain has developed around ad-hoc nomenclature for decades. While this has generally been accepted by the specialists that understand the application and purpose of unique objects, the network as a whole has suffered due to diverse intelligence being labeled with a singular term. This initial taxonomy proposal aims to address the general confusion around the cyber-threat intelligence domain by identifying the unique characteristics of cyber-threat intelligence objects. The classification of these objects should help organizations qualify the effectiveness of their intelligence consumption and analysis processes as they relate to organization risk management programs by enumerating specifically which CTI objects are currently in place and which are not.

Adam Greer, adam.greer@sans.edu

References

Common Vulnerability Scoring System v3.1: Specification Document. (n.d.). FIRST — Forum of Incident Response and Security Teams. Retrieved from www.first.org/cvss/v3.1/specification-document

Intelligence. Federal Bureau of Investigation. (2020, November 19). Retrieved from www.fbi.gov/about/leadership-and-structure/intelligence-branch

(ISC)2 Cybersecurity Workforce Study, 2019. (2019). (ISC)2. Retrieved from www.isc2.org/-/media/ISC2/Research/2019-Cybersecurity-Workforce-Study/ISC2-Cybersecurity-Workforce-Study-2019.ashx

Alberts, C., Behrens, S., Pethia, R., & Wilson, W. (1999, September). Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) Framework, Version 1.0. Software Engineering Institute. doi: 10.1184/R1/6575906.v1

Aldenderfer, MS., & Blashfield, RK. (1984). Cluster Analysis. Sage Publications.

Alter, Steven. (1977). A Taxonomy of Decision Support Systems. Sloan Management Review (pre-1986) 19, 1. ProQuest Central, 39.

Garousi, V., (et al.). (2016). The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. doi: dl.acm.org/doi/10.1145/2915970.2916008

Holst, A. (2020, March). Cybersecurity market revenues worldwide 2017-2023. Retrieved from www.statista.com/statistics/595182/worldwide-security-as-a-service-market-size/

Iivari, J. (2007). A paradigmatic analysis of information systems as a design science. Scandinavian Journal of Information Systems 19(2), 39–64. doi: https://doi.org/10.1287/isre.9.2.164

Kim, E., Gardner, D., Deshpande, S., Contu, R., Kish, D., & Canales, C. (2018, September). Forecast Analysis: Information Security, Worldwide, 2Q18 Update. Retrieved from www.gartner.com/en/documents/3889055

Kocher, P., Horn, J., Fogh, A., Genkin, D., Gruss, D., Haas, W., Hamburg, M., Lipp, M., Mangard, S., Prescher, T., Schwarz, M., & Yarom, Y. (2019). Spectre Attacks: Exploiting Speculative Execution. 40th IEEE Symposium on Security and Privacy (S\&P'19) 2019. Retrieved from spectreattack.com/spectre.pdf

Lee, Robert M. (2020, February). 2020 SANS Cyber Threat Intelligence (CTI) Survey. SANS Institute. Retrieved from www.sans.org/reading-room/whitepapers/analyst/2020-cyber-threat-intelligence-cti-survey-39395

Mavroeidis, V., & Bromander, S. (2017). Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence. 2017 European Intelligence and Security Informatics Conference (EISIC), 91-98. doi: 10.1109/EISIC.2017.20

McKnight, D., & Chernavy, N. (2001). What trust means in E-commerce customer relationships: an interdisciplinary conceptual taxonomy. International Journal of Electronic Commerce 6(2), 35–59.

Miller J., & Roth A. (1994). A taxonomy of manufacturing strategies. Management Science 40(3), 285–304.

Morgan, S. (2020). Cybercrime To Cost The World $10.5 Trillion Annually By 2025. CyberCrime Magazine. Cybersecurity Ventures. Retrieved from cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/#:~:text=Cybersecurity%20Ventures%20predicts%20global%20spending,cybersecurity%20market%20growth%20through%202021.

Nickerson, R., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. European Journal of Information Systems, 22:3, p336-359. doi: 10.1057/ejis.2012.26

Shackleford, D. (2015, February). Who's Using Cyberthreat Intelligence and How? SANS Institute. Retrieved from www.sans.org/reading-room/whitepapers/analyst/membership/35767

Sowa, J., & Zachman, J. (1992). Extending and formalizing the framework for information systems architecture. IBM Systems Journal 31(3), 590–616.

Appendix A:

	DIMENSIONS														
	Audience				Application Area					Decision Level			Implementation		Direction	
OBJECTS	Administrators	SOC Analysts	Specialists	Directors	Hardware	Software	Network	Org Sec	Landscape	Operational	Tactical	Strategic	EDP	DSS	Uni-directional	Bi-directional
IP allow/block-lists	X						X			X			X		X	
DNS allow/block-lists	X						X			X			X		X	
Update/Hotpatch Feed	X				X					X			X		X	
Hardened Baselines	X				X						X		X		X	
Support Ticket Trends	X							X			X			X		X
Vulnerability Scan Reports	X							X			X			X	X	
DNS Information Platforms		X					X				X			X	X	
IP information Platforms		X					X				X			X	X	

IOA Detections		X								X			X			X
Aggregator Platforms		X						X			X			X		X
Malware Identification Platforms		X				X				X				X		X
Anti-virus Detection Alerts		X				X				X			X		X	
IOC Detections		X						X		X				X		X
Network Traffic alerts		X						X		X				X		X
Exploit PoC disclosure			X					X		X				X	X	
Malware TTPs			X						X		X		X		X	
Threat Actor Feed			X						X		X			X		X
Aggregate vulnerability databases			X						X		X			X	X	
Control Frameworks			X					X			X			X		X
Threat Actor Report			X						X			X		X		X
Hardware End-of-Life				X	X							X		X	X	
Software end-of-life				X		X						X		X	X	

Landscape Analysis Report				X					X			X		X		X
Threat simulation Report				X				X				X		X		X
Compliance Audit Report				X				X				X		X		X