Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence

Vasileios Mavroeidis
University of Oslo
Norway
vasileim@ifi.uio.no

Siri Bromander
mnemonic
University of Oslo
Norway
siri@mnemonic.no

***Abstract*—Threat intelligence is the provision of evidence-based knowledge about existing or potential threats. Benefits of threat intelligence include improved efficiency and effectiveness in security operations in terms of detective and preventive capabilities. Successful threat intelligence within the cyber domain demands a knowledge base of threat information and an expressive way to represent this knowledge. This purpose is served by the use of taxonomies, sharing standards, and ontologies.**

This paper introduces the Cyber Threat Intelligence (CTI) model, which enables cyber defenders to explore their threat intelligence capabilities and understand their position against the ever-changing cyber threat landscape. In addition, we use our model to analyze and evaluate several existing taxonomies, sharing standards, and ontologies relevant to cyber threat intelligence. Our results show that the cyber security community lacks an ontology covering the complete spectrum of threat intelligence. To conclude, we argue the importance of developing a multi-layered cyber threat intelligence ontology based on the CTI model and the steps should be taken under consideration, which are the foundation of our future work.

***Index Terms*—cyber threat intelligence, threat information sharing, cyber security, threat intelligence ontologies, cyber attack attribution, cyber threat detection, cyber threat prevention, knowledge representation**

I. Introduction

The capabilities, persistence, and complexity of adversarial attacks in the present threat landscape result in a speed race between security analysts, incident responders, and threat actors. Coordinated cyber crime is at each peak. PwC's global economic crime survey of 2016 [1] reports that there are organizations that had suffered cybercrime losses over $5 million, and of these nearly a third reported losses in excess of $100 million. In addition, Juniper Research [2] reports that cybercrime will increase the cost of data breaches to $2.1 trillion globally by 2019; four times the estimated cost of breaches in 2015.

In the Proceedings of the European Intelligence and Security Informatics Conference (EISIC 2017), Attica, Greece, September 11-13, 2017. DOI: 10.1109/EISIC.2017.20

This research was supported by the research projects Oslo Analytics, TOCSA, and ACT funded by the Research Council of Norway.

Security analysts and incident responders need the right skills to recognize attacks before performing defense efforts. The development of adequate controls require a thorough threat analysis, but small and medium sized businesses most of the times have inadequate capabilities due to lack of skilled personnel and budget constraints.

Threat intelligence is referred to as the task of gathering 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 [1]. Threat information reported and shared between security teams is overwhelming making difficult its absorption and correlation to existed stored knowledge; as a result, threat intelligence vendors are increasingly shifting to ways of automating this process making threat analysis a viable task.

Analyzing and sharing threat data and threat information in an effective way requires common representation, standard formats and protocols for sharing, and a common understanding of the relevant concepts and terminology. A solution approach to this need is the use of artificial intelligence (AI) and particularly the use of ontologies. An ontology is a form of knowledge representation that can integrate information coming from different sources.

Working towards an ontology for cyber threat intelligence is not an easy task. Our research reports the following as the largest difficulties:

- Vaguely defined terminology leads to confusion among experts and additional work to extend or unify ontologies.
- Lack of formal standardized representation of relevant information results in strings of English prose, with no standard pattern. Standardizing well defined taxonomies can eliminate this barrier.
- Lack of coherent relationships between the different layers of abstraction in ontologies. Modular ontologies containing several sub-ontologies need sound relationships

[1] https://www.gartner.com/doc/2487216/definition-threat-intelligence

between the different data points to leverage the power of semantics and reasoning. For example, to understand the behavior and the capabilities of a threat actor the connections and relationships between pieces of information must be sound.

This article evaluates taxonomies, sharing standards, and ontologies relevant to the task of creating an ontology for use within cyber threat intelligence. Some of the ontologies potentially can aid threat intelligence but initially have been introduced to address a specific domain within cyber security. Additionally, we pinpoint the relationship between our own Cyber Threat Intelligence model (CTI), the taxonomies, the sharing standards, and the ontologies discussed, aiming to classify them in terms of expressivity. Finally, we critically discuss the shortcomings of the present cyber threat intelligence ontology approaches and we address the directions that should be followed for their advancement.

II. Methodology

This section introduces two models related to threat detection maturity and cyber threat intelligence, respectively. The two models overlap and both can meet different needs that are explained in the next two subsequent subsections. The Cyber Threat Intelligence model is the basis of the evaluation process conducted in this paper.

A. The Detection Maturity Level Model - DML

Ryan Stillions proposed the DML model in several blog postings in 2014 [3]. The model was originally used to describe the maturity of an organization in terms of their ability to consume and act upon given threat information. Threat information can include indicators of compromise, tactics techniques and procedures of an actor (TTPs), threat intelligence reports and many more. In 2016, we extended this model by adding an additional level (9) "Identity" and presented it for use in semantic representation of cyber threats [4].

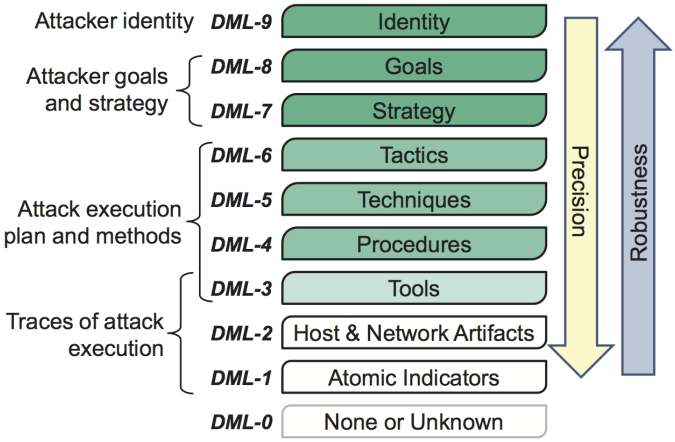

Fig. 1. Modified Detection Maturity Level Model [4] [3]

The DML model emphasizes the increasing level of abstraction in the detection of cyber attacks, where it is assumed that a security incident response team of low maturity and low skills would be able to detect attacks in terms of low level technical observations in a network, without necessarily understanding the significance of these observations. On the other hand, a security incident response team of high maturity and high skills is assumed to be able to interpret technical observations in networks in the sense that the type of attack, the attack methods used, and possibly the identity of the attacker can be determined.

Detection maturity, threat information, and threat intelligence overlap in a way that high or low detection maturity consequently can produce rich or poor threat information that can result in rich or poor threat intelligence. However, rich threat intelligence can aid the detecting and preventing capabilities of teams of low maturity by absorbing advanced threat intelligence shared from teams with higher detection capabilities.

B. The Cyber Threat Intelligence Model - CTI

For the purpose of evaluating and classifying taxonomies, sharing standards, and ontologies relevant to threat intelligence we identified the need to develop a new model that can suitably characterize threat intelligence. The Cyber Threat Intelligence model is not hierarchical like the DML model, but mainly a way to represent what types of information are needed for advanced threat intelligence and potential attack attribution. Acquisition of the Cyber Threat Intelligence model in the security operations of an organization strengthens the security posture of the organization itself by enabling advanced detective and preventive capabilities.

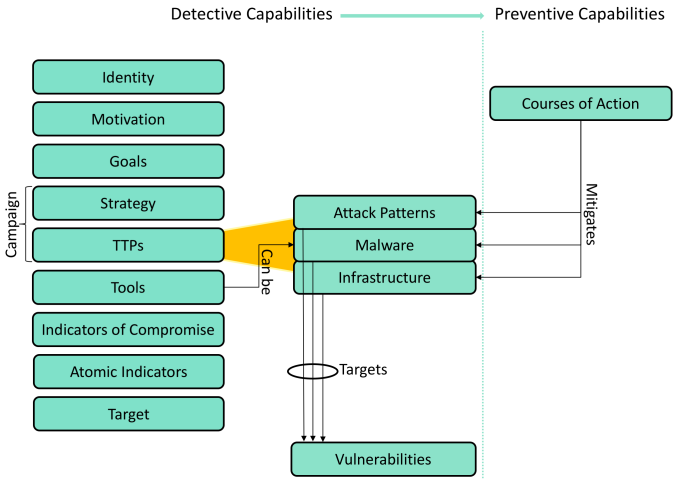

Fig. 2. Cyber Threat Intelligence Model

The remaining of the section is devoted in specifying the definitions of the elements comprising the CTI model.

Identity: The identity of a threat actor can be the name of a person, an organization, or a nation state. Sometimes, the identity can only be linked to other attacks without actual attribution or even location of their operations. However, it is important to be able to connect multiple attacks to the same

actor in order to determine any strategy, tactics, techniques, and procedures expected to be used.

Motivation: Motivation can be described as the driving force that enables actions in the pursuit of specific goals. Motivation may be derived from the benefits achieving a goal. The goals of an attacker may change, but the motivation most of the times stays the same. Knowing a threat agent's motivation narrows down which targets that agent may focus, helps defenders focus their limited defense resources on the most likely attack scenarios, as well as shapes the intensity and the persistence of an attack [5]. Examples of motivation can be ideological (human rights, ethnic etc.), military, financial and many more.

Goals: According to Fishbach and Ferguson [6] "a goal is a cognitive representation of a desired endpoint that impacts evaluations, emotions, and behaviors". A goal consists of an overall end state and the behavior objects and plans needed for attaining it. The activation of a goal guides behaviors (strategy). Depending on how the attack is organized the goal might not be known for the attack team executing the attack. The team might only receive a strategy to follow. In present cyber threat intelligence most of the times goals are described in prose. A goal can be defined as a tuple of two: (Action, Object), but work needs to be done to create a consistent taxonomy at an adequate level of detail [4]. Typical examples of goals are to "steal intellectual property", "damage infrastructure", and "embarrass competitor".

Strategy: This is a non-technical high-level description of the planned attack. There are typically multiple different ways an attacker can achieve its goals, and the strategy defines which approach the threat agent should follow. In present cyber threat intelligence, strategies are most of the times described in prose. It is our belief that the introduction of a formal taxonomy describing relationships between motives, goals, and strategies would be advantageous for the advancement of cyber threat intelligence, as well as the risk assessment processes. Part of our future research is the development of such a taxonomy.

TTPs: Tactics, techniques, and procedures characterize adversary behavior in terms of what they are doing and how they are doing it.

1) Attack Patterns: Attack Patterns are a type of TTP that describe ways the adversaries utilize to compromise targets.

2) Malware: Malware is a type of TTP and refers to a software that is inserted into a system with the intent of compromising the target in terms of confidentiality, integrity, or availability.

3) Infrastructure: Infrastructure is a type of TTP and refers to the resources of the attackers available to perform attacks. Examples of adversarial infrastructure include command and control servers, malware delivery sites, and phishing sites.

Tools: Attackers install and use tools within the victim's network. The tools often are modified so that a tool detected and analyzed in a previous security incident might be similar, but not exactly the same in new attacks. Malware is a subcategory of tools. In addition, tools might be non-malicious software (e.g., vulnerability scanners, network scanning tools) used for malicious reasons.

Indicators of Compromise: IOCs are detective in nature and describe how to recognize malicious or suspicious behavior that directly detects campaigns, TTPs, attack patterns, malware, tools, and threat actors. To create a good IOC it is desirable to combine different types of information, like atomic indicators, behavioral indicators, and computed indicators related to TTPs often referred to as "ABC" [7].

Atomic Indicators: The value of atomic indicators is limited due to the short shelf life of this type of information and can include file hashes, domain names, IPs and many more. This is the type of data and information that has the longest history in cyber threat intelligence and many threat intelligence efforts are based upon.

Target: Targets can represent organizations, companies, sectors, nations, and individuals.

Courses of Action: Courses of Action refer to measures that can be taken to prevent or respond to attacks.

C. Evaluation Criteria

In the next section of the article we cover taxonomies, sharing standards, and ontologies relevant to threat intelligence and analyze them based on the following criteria:

- Data and concepts covered based on the CTI model (Table 1).
- Connections (relationships) with other taxonomies and ontologies (Sections III, IV).
- Critical analysis of the ontologies based on the description provided in their publications or documentation, as well as their source files (Sections IV, V).

Some identified articles present ontologies which are not described in great detail and have no reference to the actual ontology (rdf/owl files), thus making their evaluation a hard task to achieve. Furthermore, some available ontologies do not offer an additional publication and most of the times not even proper documentation.

Table 1 shows concisely the results of our research conducted on taxonomies, sharing standards, and ontologies based on the CTI model.

III. Taxonomies and Sharing Standards

This section provides an overview of taxonomies and sharing standards that are used or can be used in cyber threat intelligence. We categorize them as enumerations, scoring systems, and sharing standards.

A. Enumerations

Threat Agent Library (TAL) [8] is a set of standardized definitions and descriptions to represent significant threat agents. The library does not represent individual threat actors, thus it is not intended to identify people, or investigating actual security events. The goal of TAL is to help in risk management and specifically to identify threat agents relevant to specific assets. In that way security professionals pro-actively can build defenses for specific threats. In our opinion, the defined

"hostile" threat actor types in TAL library can be used in combination to Mitre's ATT&CK taxonomy which provides a collection of known threat actors and their known tactics and techniques. The connection of the two aforementioned taxonomies would result in the introduction of a new taxonomy which classifies threat actors. An example is state actors that have government resources and their skill are considered adept.

Casey in 2015 [5] introduced a new taxonomy for cyberthreat motivations. The taxonomy identifies drivers that cause threat actors to commit illegal acts. Knowing these drivers could indicate the nature of the expected harmful actions.

Mitre's Common Vulnerabilities and Exposures (CVE) [9] dictionary provides common identifiers for publicly known information-security vulnerabilities in software packages.

NIST's National Vulnerability Database repository (NVD) [10] includes databases of security checklists, security related software flaws, mis-configurations, product names, and impact metrics (CVSS). NVD is built upon CVE and integrates CPE, as well as CWE into the scoring (impact metrics) of CVE entries.

Mitre's Common Platform Enumeration (CPE) [11] specification defines standardized machine readable methods for assigning and encoding names to IT product classes (software and hardware).

Mitre's Common Weakness Enumeration (CWE) [12] is a dictionary of software security weaknesses and vulnerabilities based in part on CVE aiming to better understand flaws in software and to propose adequate countermeasures. Their dictionary includes summaries of the attacks, the prerequisites of launching these attacks, and mitigation solutions.

Mitre's Common Attack Patterns Enumerations and Characteristics (CAPEC) [13] provides a collection of the most common techniques (methods) used in cyber attacks resulting from CWE. Like CWE, CAPEC includes summaries , attack prerequisites, and solutions (countermeasures) of the most common attack patterns.

Mitre's Adversarial Tactics, Techniques and Common Knowledge (ATT&CK) [14] provides a collection of known actors, their known tactics (10 tactic categories), and post-compromise techniques to achieve their objectives. The difference between CAPEC and ATT&CK is that the first one enumerates a range of attack patterns across the entire cyber attack life cycle whereas the latter provides comprehensive coverage across a range of post-compromise techniques. In addition to the techniques observed in ATT&CK includes tools that have been used by specific threat actors which are connected with specific techniques. Overall, these taxonomic connections help us to correlate identified indicators and TTPs to threat actor identities.

B. Scoring Systems

NIST's NVD Common Vulnerability Scoring System (CVSS) [15] is a measurement standard aiming to score vulnerabilities accurately based on their severity; as a result, CVSS enables prioritization vulnerability remediation activities.

Mitre's Common Weakness Scoring System (CWSS) [16] is part of CWE and it provides a mechanism for scoring weaknesses (CWEs) using 18 different factors. Worthy to mentioning is that Mitre's Common Weakness Risk Analysis Framework (CWRAF) can be used in conjunction with CWSS to identify the most important CWEs applying to a particular business and their deployed technologies. The difference between CVSS and CWSS is that the first one targets specific software vulnerabilities scoring, whereas the latter one targets CWE scoring.

C. Sharing Standards

A study of existing threat intelligence sharing initiatives [17] concludes that structured threat information eXpression (STIX) is currently the most used standard for sharing structured threat information. STIX [18] is an expressive, flexible, and extensible representation language used to communicate an overall piece of threat information. STIX architecture is comprised of several cyber threat informations such as cyber observables, indicators, incidents, adversaries tactics, techniques, procedures, exploit targets, courses of action, cyber attack campaigns, and threat actors. Furthermore, STIX was recently redesigned and as a result, omits some of the objects and properties defined in the first version. The objects chosen for inclusion in the second version represent a minimally viable product (MVP) that fulfills basic consumer and producer requirements for CTI sharing. Both standards can be used and adapted based on an organization's needs. It is worth pointing out that MITRE offers additionally Malware Attribute Enumeration and Characterization (MAEC) [19], which is a very expressive malware sharing language for encoding and communicating high-fidelity information about malware based upon attributes such as behaviors, artifacts, and attack patterns. MAEC can be integrated in STIX or used as a standalone.

OpenIOC, originally developed by Mandiant, is an extensible XML schema that enables you to describe the technical characteristics that identify a known threat, an attacker's methodology, or other evidence of compromise. The types of information covered directly by OpenIOC are derived mainly by low level atomic indicators, comprising indicators of compromise, thus covering the IOC category of the CTI model.

IV. Ontologies

Since the work of Blanco et al. [20] in 2008, we have not found any overviews of existing ontologies within the cyber security domain. The authors remark that the scientific community has not accomplished a general security ontology because most of the works are focused on specific domains or the semantic web. The same conclusion was drawn by Fenz and Ekelhart [21]. Additionally, Blanco et al. [20] emphasize the complication of combining their identified ontologies due to the non-common interpretation and different terms applied

for similar concepts in different ontologies. Our study confirms the same almost 10 years after the study of Blanco et al. [20].

Cyber threat information is a small subsection of the information relevant to cyber security and the full security domain. While several ontologies relevant to cyber security and security analytics exist, few ontologies related to threat information and threat intelligence can be identified. We have listed the ontologies discovered relevant to cyber threat intelligence and some more general security ontologies that look promising, at least conceptually, to be taken under consideration when working towards a full cyber threat intelligence ontology. In addition, for many ontologies, relation to specific CTI categories is a tough assignment due to their limitation of being described at a very high level. For most of the ontologies we were unable to find the relevant rdf/owl files even though many of them are called "open-source" by the authors. The ontologies analyzed hereafter are listed chronologically based on the publication date.

Stefan Fenz and Andreas Ekelhat [21] described an information security ontology that can be used to support a broad range of information security risk management methodologies. The high level concepts of the ontology are based on the security relationship model described in the National Institute of Standards and Technology Special Publication 800-12 and is comprised of the threat, vulnerability, control, attribute, and rating concepts to represent the information security domain knowledge. In addition, concepts such as asset, organization, and person are necessary to formally describe organizations and their assets. Lastly, the concept of location is integrated combined with a probability rating concept to interrelate location and threat information in order to assign priory threat probabilities. Like most of the works the authors have difficulties to connect unambiguous concepts from different standards such as the distinction between threats and vulnerabilities.

Wang and Guo [22] proposed an ontology for vulnerability management and analysis (OVM) populated with all existing vulnerabilities in NVD. The basis of the ontology is built on the results of CVE and its related standards such as CWE, CPE, CVSS, and CAPEC. OVM captures the relationships between the following concepts which constitute the top level of the ontology; vulnerability, introduction phase (software development life cycle - time periods during which the vulnerability can be introduced), active location (location of the software where the flaw manifests), IT product, IT vendor, product category (such as web browsers, application servers, etc.), attack (integration of CAPEC), attack intent, attack method, attacker (human being or software agent), consequence, and countermeasure.

Obrst et al. [23] suggested a methodology for creating an ontology based on already well-defined ontologies that can be used as modular sub-ontologies. In addition, they remark the usefulness of existing schemas, dictionaries, glossaries, and standards as a form of knowledge acquisition of the domain by identifying and analyzing entities, relationships, properties, attributes, and range of values that can be used in defining an ontology. Their suggested ontology is based on the diamond model of malicious activity [24], which expresses the relationships between an adversary (actor), the capabilities of the adversary, the infrastructure or resources the adversary utilizes, and the target of the adversary (victim). The authors state that they developed first the aspects of infrastructure and capabilities, but they are still not in the level of detail they desire. In addition, their current ontology is focused on malware and some preliminary aspects of the diamond model.

A good argumentation for transitioning from taxonomies to ontologies for intrusion detection was made in 2003, by Undercoffer et al. [25]. They suggested an ontology that would enable distributed anomaly-based host IDS sensors to contribute to a common knowledge-base, which again would enable them to quicker identify a possible attack.

Based on this, More et al. [26] in 2012, suggested to build a knowledge-base with reasoning capabilities to take advantage of an extended variety of heterogeneous data sources, to be able to identify threats and vulnerabilities. Their data sources suggest that data retrieved and included in the ontology is within the atomic indicators category of the CTI model.

Oltramari et al. [27] proposed a three layer cyber security ontology named "CRATELO" aiming to improve the situational awareness of security analysts, resulting to optimal operational decisions through semantic representation. Following the methodology of [23], the authors build upon existing ontologies and expand them. Specifically, CRATELO includes the top level ontology DOLCE-SPRAY extended with a security related - middle level ontology (SECCO) capable to capture details of domain specific scenarios such as threat, vulnerability, attack, countermeasure, and asset. The low level sub-ontology, cyber operations (OSCO), is the extension of the middle level ontology.

Gregio et al. [28] suggested an ontology to address the detection of modern complex malware families whose infections involve sets of multiple exploit methods. To achieve this, they created a hierarchy of main behaviors each one of them consisting of a set of suspicious activities. Then they proposed an ontology that models the knowledge on malware behavior. They state that a given program behaves suspiciously if it presents one or more of the six events (main behaviors) described below which consist of several characteristics. The events are attack launching, evasion, remote control, self-defense, stealing, and subversion. When new set of process actions with malicious behaviors appear (input from "transformed" log files), the ontology can be inferred to see if an instance of suspicious execution is linked to a malware sample.

Salem and Wacek [29] designed a data extraction tool called TAPIO (Targeted Attack Premonition using Integrated Operational data) which is specialized in extracting data (natural language processing) and automatically map them into a fully linked semantic graph accessible in real time. Part of TAPIO is a cyber security ontology going by the name Integrated Cyber Analysis System (ICAS) that ingests extracted data (logs and events) from several sources to provide relationships across an enterprise network. The tool aims to help incident response teams in connecting and correlating events and actions into an

ontology for automatic interpretation. ICAS is a collection of 30 sub-ontologies specializing in specific conceptual areas as part of host based and network based conceptual models.

Iannacone et al. [30] described their STUCCO ontology, which is developed to work on top of a knowledge graph database. The STUCCO ontology design is based upon scenarios of use by both human and automated users and incorporates data from 13 different structured data sources with different format. The data included in the current STUCCO ontology fall into the categories identity, TTPs, tools, and atomic indicators of the CTI model. Their future work included extending the ontology to support STIX.

Gregio, Bonacin, de Marchi, Nabuco, and de Geus [31] expanded the work of Gregio et al. [28] and introduced the malicious behavior ontology (MBO). MBO is capable of detecting modern complex malware families whose infections involve sets of multiple exploit methods, by applying SWRL rules to the ontology for inferencing. In addition, these rules also apply metrics to specify whether a program is behaving maliciously or not and specifically, how suspicious the execution of a program is. The authors state that their model is able to detect unknown malicious programs even in cases where traditional security mechanisms like antivirus are not, by performing automatic inference of suspicious executions in monitored target systems. However, the current state of the ontology has some limitations such as performance issues, cannot detect malware in real time, and false positives and negatives.Based on its operation MBO can provide useful indicators of compromise for malware.

Fusun et al. suggested ontologies for quantifying attack surfaces [32]. Their Attack Surface Reasoning (ASR) gives a cyber defender the possibility to explore trade-offs between cost and security when deciding on composition of their cyber defense. Ontologies created include those of attacks, systems, defenses, missions and metrics. ASR is mainly modeled after the Microsoft STRIDE [33] threat classification framework, which categorizes attack steps into 6 categories and is to the extent of our knowledge not the preferred framework within threat intelligence community due to its lack of details. In comparison, CAPEC and CPE have around 500 and 1000 ”categories” respectively.

As part of their study on using security metrics for security modeling, Pendelton et al. suggested the Security Metric Ontology [34]. Their ontology includes four sub-ontologies; vulnerability, attack, situations and defense mechanisms, and describes the relationship between them. The terminology used is somewhat different than that of known taxonomies, and their aim at modeling metrics is more prominent than that of analysis and reasoning. Their ontology is published on GitHub[2].

Unified Cybersecurity Ontology was suggested by Syed et al. [35] in 2016. It serves as a backbone for linking cyber security and other relevant ontologies. There are mappings to aspects of STIX, and references to CVE, CCE, CVSS, CAPEC, STUCCO and KillChain. The mappings are loosely connected at a very high level. It is worthy to note that they do not make use of OWL constructs which reduces the reasoning capabilities of the ontology. In addition, their use of domain and range restrictions would result in faulty classification when used with a reasoner. Their ontology is published on GitHub[3].

Unified Cyber Ontology has been introduced on GitHub[4], without any academic publications to date and no actual rdf/owl files yet. Their model ontology is however interesting as it originates from the creators of STIX, which is currently the most used format for sharing threat intelligence [17]. The content of their work is driven primarily from the initial base requirements of expressing cyber investigation information and is the product of input from the Cyber-investigation Analysis Standard Expression (CASE) community (CASE)[5].

Without any publication we find the Cyber Intelligence Ontology (CIO), published only on GitHub[6] to be relevant. This GitHub repository includes most of the mentioned taxonomies and sharing standards in this article, extended and transformed in OWL. The limitation of those ontologies is that they are not connected or unified. For the aforementioned reason we do not classify CIO, since details can be found by checking the relevant taxonomies and sharing standards described in this paper.

V. Discussion

Threat intelligence demands great attention from any organization entailing advanced cyber-threat detective and preventive capabilities. The goal of threat intelligence is to gain rich evidence that can aid decision making, thus the maturity, the skills, and the information sources of a security team define their capability to produce accurate and actionable threat information [39] [40]. Security teams of any maturity and skills can benefit from information sharing activities allowing someone’s detection to become another’s prevention. By exchanging threat information, organizations can leverage the collective knowledge to get a better understanding of threats an organization might face and consequently, improve their security posture.

Relationships and reasoning: For leveraging the power of ontologies and description logic all the abstraction layers of the CTI model need to be introduced and taken under consideration by formalizing their relationships. In the ontologies evaluated we concluded that lack of OWL constraints is a common phenomenon. Constraints (restrictions) are what makes OWL powerful and enable its reasoning capabilities by inferring information from asserted information. In addition, most of the ontologies explicitly target specific sub-domains of threat intelligence, thus limiting the decision making process in the presence of an observed threat. Successful integration, operation, and advanced reasoning capabilities require existing taxonomies, standards, and vocabularies interconnected in

[2]https://github.com/marcusp46/security-metrics-ontology

[3]https://github.com/Ebiquity/Unified-Cybersecurity-Ontology

[4]https://github.com/ucoProject/uco

[5]https://github.com/casework/case

[6]https://github.com/daedafusion/cyber-ontology

TABLE I
CTI EVALUATION: TAXONOMIES, SHARING STANDARDS, AND ONTOLOGIES

		Identity	Motivation	Goal	Strategy	TTP	Tool	IOC	Atomic Indicator	Target	COA
Taxonomies	TAL [8]	*									
	Threat Agent Motivation [5]	*	*								
	CVE [9]								*		
	NVD [10]								*		
	CPE [11]								*		
	CWE [12]					*			*		*
	CAPEC [13]					*		*			*
	ATT&CK [14]	*				*	*				
	CVSS [15]								*		
	CWSS [16]								*		
Sharing Standards	STIX 1 [18]	*	*	* (Intended Effect:taxonomy)	*	*	*	*	*	*	*
	STIX 2 [36]	*	*	* (Objectives:string)	*	*	*	*	*	*	*
	MAEC [19]							*			
	OpenIOC [37]					*	*	*	*		
Ontologies	Fenz & Ekelhat (2009) [21]								*		
	Wang & Guo (2009) - OVM [22]					*			*		*
	Orbst et al. (2012) [23]	*					*		*	*	
	More et al. (2012) [26]					*			*		
	Oltramari et al. (2014) - CRATELO [27]	*				*			*	*	
	Gregio et al. (2014) [28]						* (malware)		*		
	Salem & Wacek (2015) - ICAS [29]					*			*		
	Iannacone et al. (2015) - STUCCO [30]	*				*	*		*		
	Gregio et al. (2016) - MBO [31]						* (malware)	*	* (it may provide)		
	Fusun et al. (2015) - ASR [32]					*			*	*	
	Pendelton et al. (2016) - Security Metrics Ontology [34]					*					*
	Syed et al. (2016) - UCO [35]	*	*	*	*	*	*	*	*		*
	Unified Cyber Ontology (2016) - UCO [38]	*	*	*	*	*	*	*	*	*	*

addition to human domain expertise to create an ontology that reasons between all the abstraction layers. Furthermore, we cannot ignore the lack of interconnection between taxonomies related to motivations, goals, and strategies of the attackers which can be used multi-purposely. The importance of these taxonomies can be seen in cases that we want to identify which threat actors target particular sectors, and ways of infiltration often used based on their motives, goals, strategies, and TTPs.

Knowledge collection: Much of the knowledge used by most of skilled analysts today is residing only in their heads. If we manage to model this knowledge and express it in an ontology, not only more analysts would be able to consume this type of intelligence, but the analytics would be executed in a consistent way, contradicting the "confirmation bias" often referred to in an investigation. To be able to express the knowledge of skilled analysts through an ontology, we need to gather their knowledge without expecting them to have prior knowledge of ontologies. We suggest doing this in a iterative way, conducting interviews to pinpoint the process, data sources and actual reasoning points used by highly skilled personnel.

Attribution: Attribution of attacks is the most important element of threat intelligence both for direct recipients and general public. To be able to attribute an attack, evidence of operations is needed that can be linked to an attacker. This entails data and information from different categories of the CTI model. Relating the data points to each other is the task of data enrichment.

The current most common bases for attribution claims include [41] timestamps in executable files; strings, debug paths, and metadata in binary sources such as malware and infected documents; reuse of infrastructure and back-end connections; malware families; code reuse; reused passwords (email accounts, encrypted pieces of code); exploits (0-days); targets (states, secret agencies, etc.).

According to Bartholomew and Guerreri-Saade [41] for the aforementioned bases only sloppy actors or careless operators will provide more data than they should, like debug paths and language strings, or reuse infrastructure from previous attacks. Rid and Buchanan [39] agree with the most common bases for attribution claims but also state that language indicators remain a worthy part of the attribution process. The authors [39] additionally remark that the attackers often re-use software to accomplish basic tasks in their operations for efficiency reasons.

Trust and uncertainty: Attribution is related to uncertainty since intermediate to advanced threat actors are aware of attribution methods and adapt several masquerading techniques. In addition, we find different degrees of knowledge among those sharing threat intelligence enabling sharing of possibly faulty or inaccurate threat intelligence. Added to that, comes the absence of having standardized requirements related to the quality of evidence before shared, which in many cases creates just as large amplifications as the actual threat itself.

For the reasons described we need to take into consideration the level of certainty related to a single piece of information

and the level of trust we have in a given source, that being human or computer device. To address this issue we suggest the use of subjective logic [42] in modeling trust of sources, and confidence in pieces of intelligence that can result in expanding the situational awareness of a security analyst.

VI. Conclusion

Our study concludes that there is not any existing ontology readily available for use within cyber threat intelligence. The main shortcoming is the lack of expressiveness resulting from their poor development and the fact that none of them covers all the relevant data and information (abstraction layers) needed for effective cyber threat intelligence. We suggest several tasks that need addressing in order to create a multi-layered cyber threat intelligence ontology. First, formal terminology (definitions) and vocabularies should be described. Second, all the abstraction layers of the cyber threat intelligence model should be included and expressed properly in the ontology. Third, knowledge coming from domain expertise in a structured way should be gathered and formally represented in the ontology to facilitate advanced reasoning based on relationships between data. Fourth, constraints should be defined and constructs should be used in the ontology enabling the reasoning capabilities lying within the OWL language. Finally, the use of subjective logic to model trust in sources and confidence in information.

References

[1] K. McConkey, "Cybercrime: A Boundless Threat," https://www.pwc.com/gx/en/services/advisory/forensics/economic-crime-survey/cybercrime.html.

[2] S. Smith, "Cybercrime will Cost Businesses over $2 Trillion by 2019," https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion.

[3] R. Stillions, "The DML Model," http://ryanstillions.blogspot.com/2014/04/the-dml-model_21.html.

[4] S. Bromander, A. Jøsang, and M. Eian, "Semantic Cyberthreat Modelling," in *STIDS*, 2016, pp. 74–78.

[5] T. Casey, "Understanding cyber threat motivations to improve defense," *Intel White Paper*, 2015.

[6] A. Fishbach and M. J. Ferguson, "The goal construct in social psychology." 2007.

[7] SANS, "Security Intelligence: Attacking the Cyber Kill Chain," https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain/.

[8] T. Casey , "Threat Agent Library Helps Identify Information Security Risks," *Intel White Paper, September*, 2007.

[9] MITRE, "Common Vulnerabilities and Exposures," https://cve.mitre.org.

[10] NIST, "National Vulnerability Database," https://nvd.nist.gov/.

[11] Mitre, "Common Platform Enumeration," https://cpe.mitre.org/specification/.

[12] MITRE, "Common Weakness Enumeration," https://cwe.mitre.org.

[13] MITRE, "Common Attack Pattern Enumeration and Classification," https://capec.mitre.org/.

[14] MITRE, "Adversarial Tactics, Techniques and Common Knowledge," https://attack.mitre.org/.

[15] NIST, "Common Vulnerability Scoring System," https://nvd.nist.gov/vuln-metrics/cvss.

[16] MITRE, "Common Weakness Scoring System," https://cwe.mitre.org/cwss/cwss_v1.0.1.html.

[17] C. Sauerwein, C. Sillaber, A. Mussmann, and R. Breu, "Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives," 2017.

[18] S. Barnum, "Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX ™)," *MITRE Corporation*, vol. 11, 2012.

[19] Mitre, "Malware Attribute Enumeration and Characterization," https://maec.mitre.org.

[20] C. Blanco, J. Lasheras, R. Valencia-García, E. Fernández-Medina, A. Toval, and M. Piattini, "A Systematic Review and Comparison of Security Ontologies," in *Availability, Reliability and Security, 2008. ARES 08. Third International Conference on*. Ieee, 2008, pp. 813–820.

[21] S. Fenz and A. Ekelhart, "Formalizing Information Security Knowledge," in *Proceedings of the 4th international Symposium on information, Computer, and Communications Security*. ACM, 2009, pp. 183–194.

[22] J. A. Wang and M. Guo, "OVM: An Ontology for Vulnerability Management," in *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies*. ACM, 2009, p. 34.

[23] L. Obrst, P. Chase, and R. Markeloff, "Developing an Ontology of the Cyber Security Domain," in *STIDS*, 2012, pp. 49–56.

[24] S. Caltagirone, A. Pendergast, and C. Betz, "The diamond model of intrusion analysis," DTIC Document, Tech. Rep., 2013.

[25] J. Undercoffer, A. Joshi, and J. Pinkston, "Modeling Computer Attacks: An Ontology for Intrusion Detection," in *International Workshop on Recent Advances in Intrusion Detection*. Springer, 2003, pp. 113–135.

[26] S. More, M. Matthews, A. Joshi, and T. Finin, "A Knowledge-Based Approach to Intrusion Detection Modeling," in *Security and Privacy Workshops (SPW), 2012 IEEE Symposium on*. IEEE, 2012, pp. 75–81.

[27] A. Oltramari, L. F. Cranor, R. J. Walls, and P. D. McDaniel, "Building an Ontology of Cyber Security," in *STIDS*. Citeseer, 2014, pp. 54–61.

[28] A. Grégio, R. Bonacin, O. Nabuco, V. M. Afonso, P. L. De Geus, and M. Jino, "Ontology for Malware Behavior: a Core Model Proposal," in *WETICE Conference (WETICE), 2014 IEEE 23rd International*. IEEE, 2014, pp. 453–458.

[29] M. B. Salem and C. Wacek, "Enabling New Technologies for Cyber Security Defense with the ICAS Cyber Security Ontology," in *STIDS*, 2015, pp. 42–49.

[30] M. Iannacone, S. Bohn, G. Nakamura, J. Gerth, K. Huffer, R. Bridges, E. Ferragut, and J. Goodall, "Developing an Ontology for Cyber Security Knowledge Graphs," in *Proceedings of the 10th Annual Cyber and Information Security Research Conference*. ACM, 2015, p. 12.

[31] A. Grégio, R. Bonacin, A. C. de Marchi, O. F. Nabuco, and P. L. de Geus, "An Ontology of Suspicious Software Behavior," *Applied Ontology*, vol. 11, no. 1, pp. 29–49, 2016.

[32] M. B. Fusun, A. S. Yaman, T. Marco, E. Carvalho, and C. N. Paltzer, "Using Ontologies to Quantify Attack Surfaces."

[33] A. Shostack, *Threat Modeling: Designing for Security*. John Wiley & Sons, 2014.

[34] M. Pendleton, R. Garcia-Lebron, J.-H. Cho, and S. Xu, "A Survey on Systems Security Metrics," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, p. 62, 2016.

[35] Z. Syed, A. Padia, M. L. Mathews, T. Finin, and A. Joshi, "UCO: A Unified Cybersecurity Ontology," in *Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security*. AAAI Press, 2016.

[36] OASIS CTI TC, "Structured Threat Information Expression (STIX ™) 2.0," https://oasis-open.github.io/cti-documentation/, 2017.

[37] Mandiant Corporation, "Sophisticated Indicators for the Modern Threat Landscape: An Introduction to OpenIOC," http://www.openioc.org/resources/An_Introduction_to_OpenIOC.pdf, 2013.

[38] "Unified Cyber Ontology," https://github.com/ucoProject/uco, 2016.

[39] T. Rid and B. Buchanan, "Attributing Cyber Attacks," *Journal of Strategic Studies*, vol. 38, no. 1-2, pp. 4–37, 2015.

[40] C. Johnson, L. Badger, D. Waltermire, J. Snyder, and C. Skorupka, "Guide to cyber threat information sharing," *NIST Special Publication*, vol. 800, p. 150, 2016.

[41] B. Bartholomew and J. A. Guerrero-Saade, "Wave your False Flags! Deception Tactics Muddying Attribution in Targeted Attacks," in *Virus Bulletin Conference*, 2016.

[42] A. Jøsang, "Artificial Reasoning with Subjective Logic," in *Proceedings of the second Australian workshop on commonsense reasoning*, vol. 48. Perth:[sn], 1997, p. 34.