Building an Ontology of Cyber Security

Alessandro Oltramari and Lorrie Faith Cranor
CyLab, Carnegie Mellon University
Pittsburgh, USA

Robert J. Walls and Patrick McDaniel
Department of Computer Science
Pennsylvania State University
University Park, USA

Abstract—Situation awareness depends on a reliable perception of the environment and comprehension of its semantic structures. In this respect, the cyberspace presents a unique challenge to the situation awareness of users and analysts, since it is a unique combination of human and machine elements, whose complex interactions occur in a global communication network. Accordingly, we outline the underpinnings of an ontology of secure operations in cyberspace. We present the basic architecture of the ontology and provide a modeling example. We make the case for adopting a rigorous semantic model of cyber security to overcome the current limits of the state of the art.

Keywords—cyber security, ontology, situation awareness, ontology patterns.

I. INTRODUCTION

As disclosed by a recent report\(^1\), there has been half a billion cyber security breaches in the first semester of 2014, matching the record set across the entire precedent year. In general, this alarming trend should not surprise when we consider that the bedrock of the Internet is a technological infrastructure built almost 35 years ago for trusted military communications and not for data exchange in the wild (see [1], p.58). The picture gets even worse when considering that the ability to grasp the risk and threats associated with computer networks is averagely poor: recent surveys have actually shown that 65% of the companies victim of intrusion and information theft are usually notified by third parties and that the detection process usually takes 13 months (e.g., see [2], p.10).

Though not exhaustive, such rough statistics at least suggests that if the inadequacy of the technological infrastructure is a key aspect to explain the vulnerabilities of networked computer systems, the human factor also plays a central role. As proposed in [3], to improve situation awareness of users and security operators, a shift of focus from system to environment level is highly necessary when modeling cyber scenarios: to this end, a full-fledged science of cyber security needs to be founded, whose core principle is cognizing the cyberspace as a hybrid framework of interaction between humans and computers, where security and privacy policies play also a crucial role. As stated by [4], this cognizance depends on both a reliable perception of the elements of an environment and, most importantly for our work, on the explicit representation of their semantics. Accordingly, the current article presents the underpinnings of an ontology of secure cyber operations: by and large, the concepts and the relationships that structure this semantic model are peculiar to the domain. That is, notions that are suitable for representing security in the physical world cannot be directly transferred to the cyber environment (e.g., attack attribution [5]). We build upon existing ontologies, expanding them to support novel use cases as needed\(^2\). Our goal is to use the proposed ontology as basis for improving the situation awareness of cyber defenders, allowing them to make optimal operational decisions given the current environment.

The remaining of the paper is organized as follows: section II will briefly make the case for the adoption of ontologies in the cyber security realm; section III will outline the structure of ‘CRATELO’, a Three Level Ontology for the Cyber Security Research Alliance program funded by ARL\(^3\), and describe a simple cyber scenario modeled by means of our approach; finally section IV draws preliminary conclusion and outlines an agenda for future research.

II. RELATED WORK

Developing cyber security ontologies is a critical step in the transformation of cyber security to a science. In 2010, the DoD sponsored a study to examine the theory and practice of cyber security, and evaluate whether there are underlying fundamental principles that would make it possible to adopt a more scientific approach. The study team concluded that:

The most important attributes would be the construction of a common language and a set of basic concepts about which the security community can develop a shared understanding. A common language and agreed-upon experimental protocols will facilitate the testing of hypotheses and validation of concepts [6].

The need for controlled vocabularies and ontologies to make progress toward a science of cyber security is recognized in [7] and [8] as well. In this domain, ontologies would include, among other things, the classification of cyber attacks, cyber incidents, malicious and impacted software programs. From our point of view, where the human component of cyber security is crucial, the analysis needs to be expanded to the different roles that attackers, users, defenders and policies play in the context of cyber security, the different tasks that the members of a team are assigned to by the team leader, and the knowledge, skills and abilities needed to fulfill them.

\(^1\) [https://www.riskbasedsecurity.com/reports/2014-MidYearDataBreachQuickView.pdf]

\(^2\) For instance, exploiting material available in this portal: [http://militaryontology.com/cyber-security-ontology.html]

\(^3\) [http://cra.psu.edu/]
There has been little work on ontologies for cyber security and cyber warfare. Within a broader paper, there is a brief discussion of an ontology for DDoS attacks [9] and a general ontology for cyber warfare is discussed in [10]. To the best of our knowledge, Obrost and colleagues [11] provide the most comprehensive description of a cyber ontology architecture, whose vision has actually inspired the work presented in this paper (the scale of the project and its difficulties are also discussed by Dipert in [8]). By and large, efforts that have been made toward developing ontologies of cyber security, even when expressed in OWL, RDF or other XML-based formats, typically do not utilize existing military domain or middle-level ontologies such UCORE-SL. With regard to human users and human computer interaction, the most important step in understanding a complex new domain involves producing accessible terminological definitions and classifications of entities and phenomena, as stressed in [7]. Discussions of cyber warfare and cyber security often begin with the difficulties created by misused terminology (such as characterizing cyber espionage as an attack): in this regard, the Joint Chiefs of Staff created a list of cyber term definitions that has been further developed and improved in a classified version. None of these definitions, however, are structured as an ontology. Likewise, various agencies and corporations (NIST, MITRE, Verizon) have formulated enumerations of types of malware, vulnerabilities, and exploits. In particular MITRE, which has been very active in this field, maintains two dictionaries, namely CVE (Common Vulnerabilities and Exposures) and CWE (Common Weakness Enumeration), a classification of attack patterns (CAPEC - Common Attack Pattern Enumeration and Classification), and an XML-structured language to represent cyber threat information (STIX - Structure Threat Information Expression). Regardless of the essential value of these resources, without a “shared semantics” the sprawling definitions they contain are hard to maintain and port into machine-readable formats.

III. A THREE-LEVEL ONTOLOGY FOR THE CYBER-SECURITY RESEARCH ALLIANCE

Top-level ontologies capture generic characteristics of world entities, such as spatial and temporal dimensions, morphology (e.g., parts, edges, sides), qualities (e.g., color, volume, electric charge), etc; because of their inherent generality, they are not suited to model contextual aspects. Nevertheless, it’s good practice to describe the fine-grained concepts that constitute a domain-level ontology in terms of foundational (or top-level) categories, adding core (or middle-level) notions to fill contingent conceptual gaps. For instance, an ontology of mineralogy should include notions like “basaltic rock”, “texture” and “metamorphic reaction”. In order to describe the meaning of those specific concepts, high-level categories such as “object”, “quality” and “process” must be employed; the ontology should also define an intermediate notion like “metamorphism”, which is common across domains (biology, chemistry, computer science, architecture, etc.), to explain how the different phases, end products, and features of metamorphic reactions are bound together.

Our ontology of cyber security makes no exceptions to the tripartite layering described above: in particular, CRATELO is an ontological framework constituted of a domain ontology of cyber operations (OSCO), designed on the basis of DOLCE top ontology, extended with a middle-level security-related ontology (SECCO), and encoded in OWL-DL. The three levels of CRATELO (schematized in figure 1) currently include 207 classes and 131 relationships (divided into 116 object properties and 15 datatype properties). The expressivity of the ontology is SRIQ, a decidable extension of the description logic SHIN (see [12] for more details).

A. Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)

DOLCE is part of a library of foundational ontologies for the Semantic Web developed under the WonderWeb project (funded by the European Commission). As reflected in the acronym, DOLCE holds a cognitive bias, i.e. aiming at capturing the conceptual primitives underlying natural language and commonsense reasoning. In order to reduce the complexity of the axiomatisation, in the current work we adopt DOLCE-SPRAY, a simplified version of DOLCE. The root of the hierarchy of DOLCE-SPRAY is ENTITY, which is defined as the class of anything that is identifiable as an object of experience or thought. The first relevant distinction is among CONCRETE ENTITY, i.e. objects located in definite spatial regions, and ABSTRACT-ENTITY, whose instances don’t have spatial properties. CONCRETE-ENTITY is further split in CONTINUANT and OCCURRENT, namely entities without inherent temporal parts (e.g. artifacts, animals, substances) and entities with inherent temporal parts (e.g. events, actions, states) respectively. The basic ontological distinctions are maintained: DOLCE’S ENDURANT and PERSISTENT match DOLCE-SPRAY’S CONTINUANT and OCCURRENT; the core difference comes from merging ABSTRACT and NON-PHYSICAL-ENDURANT categories into DOLCE-SPRAY’S ABSTRACT-ENTITY. Among subtypes of

---

4 http://www.slideshare.net/BarrySmith3/universal-core-semantic-layer-uorcsl
http://publicintelligence.net/dod-joint-cyber-terms/
http://www.nist.gov/
http://www.mitre.org/
http://www.verizon.com/
https://cve.mitre.org/
http://cwe.mitre.org/
https://capec.mitre.org/
https://stix.mitre.org/language/version1.1.1/

11 http://wonderweb.man.ac.uk/

13 Categories are indicated in small caps; relationships in italics. Presenting the axiomatisation of DOLCE-SPRAY is out of scope in this paper.
this category, CHARACTERIZATION is defined as a mapping of n-uples of individuals to truth-values. Individuals belonging to CHARACTERIZATION can be regarded to as ‘reified concepts’ (e.g. "manufactured object"), and the irreflexive, antisymmetric relation CHARACTERIZE associates them with the objects they denote ("a collection of vintage shoes"). Among the relevant sub-types of CHARACTERIZATION we can find: ROLE, i.e. the classification of an entity according to a given context or perspective; PLAN, namely the generic description of an action (such as "the disassembly of a 9mm"); TASK, that is a representation of the specific steps that are needed to execute an ACTION according to a PLAN (e.g. "removing the magazine", "pull back the slide"); REQUIREMENT, whose instances can be seen as the preconditions that need to be satisfied as part of a PLAN (e.g. "the weapon must be clear before proceeding"). The branch of the DOLCE-SPRAY rooted on CHARACTERIZATION distills the extensions introduced in [15]. An overview of DOLCE-SPRAY’s backbone taxonomy is represented in figure 2.

B. Security Core Ontology (SECCO)

In order to extend DOLCE-SPRAY for characterizing security, we designed a minimal set of ontological concepts. In particular, an entity is a THREAT $\phi$ for an ASSET $\alpha$ valued by a STAKEHOLDER $\sigma$ and protected by a DEFENDER $\delta$, if and only if $\phi$ can be used by an ATTACKER $\kappa$ to exploit a VULNERABILITY $\omega$ of $\alpha$ in an ATTACK $\tau$. To avoid $\phi$, a specific collection of SECURITY-REQUIREMENTS $\pi$ need to be satisfied by $\alpha$. When $\tau$ is launched, $\delta$ has to perform a suitable OPERATION $\circ$ and deploy COUNTERMEASURES $\gamma$ to proactively defend $\alpha$; $\circ$ is carried out on the basis of a MISSION-PLAN $\pi$ whose sequence of MISSION-TASKS $\xi$ is to be executed in $\circ$. In order to delineate $\pi$, $\delta$ needs also to preemptively assess the RISK $\rho$ associated to $\tau$ (datatype properties can be used to represent $\rho$ as a parameterization of the expected losses, probabilities of attack, etc.).

The list of class-inclusions below (1-6) denotes the alignment between SECCO and DOLCE-SPRAY’s main categories. Note that the multiple inheritance in (1-3) is based on ‘rigidity’, a meta-class defined in [16]: in summary, we can say that ‘agent’ is a rigid class because every entity that is classified by this notion fulfills some necessary requisites; on the contrary, ‘role’ is an anti-rigid class because an entity can be conceived as an attacker only on some situations. To put it in other words: we, as persons, will be always agents, no matter what we do; but only if we attack someone or something, we can be considered both agents and attackers.

...
analysis of security where the notion of FAULT is introduced to denote an interruption of the services delivered by a given system in the environment [20]. Though a middle-level ontology of security can be optionally extended beyond the minimal set of concepts outlined at the beginning of this section, what is original in our approach is the characterization of SECCO’s security primitives using DOLCE-SPRAY.

C. Ontologies of Secure Cyber Operations (OSCO)

One of the major cyber security problems for government and corporations is the widespread “operational chaos” experienced by analysts, as Michael Susong has recently called the phenomenon of “having too many alarms (false positives) in a network, not enough trained people to deal with them, and a consequent poor prioritization of risks and countermeasures” [18]. In this regard, the objective of an ontology of cyber security is to shape that chaos into a framework of meaningful and reusable chunks of knowledge, turning the operational disarray into a systematic model by means of which analysts can increase their situation awareness of cyber operations. As mentioned before (see section 1), the key to this augmented cognition relies on a consistent assessment of the situation and on a comprehensive understanding of its elements at the semantic level. But how is a cyber operation usually defined? In a document released in 2010, the Joint Chiefs of Staff describes a “cyberspace operation” as the “employment of cyber capabilities where the primary purpose is to achieve objectives in or through cyberspace. Such operations include computer network operations and activities to operate and defend the Global Information Grid” [21]. Drawing on this broad definition, and utilizing the concepts introduced in sections A and B, we represent a CYBER-OPERATION $\psi$ as an OPERATION executed by a CYBER-OPERATOR $\phi$, who plays the role of DEFENDER in the context of a CYBER-ATTACK $\theta$ (7-8). For the sake of this article, we will limit ourselves to provide an example of a cyber operation with CRATELO, since the development of a full-scale domain ontology is currently underway within our project.

\[
\begin{align*}
CYBER - OPERATION & \sqsubseteq OPERATION \hspace{2cm} (7) \\
CYBER - OPERATOR & \sqsubseteq DEFENDER \hspace{2cm} (8) \\
CYBER - ATTACK & \sqsubseteq ATTACK \hspace{2cm} (9)
\end{align*}
\]

1) Modeling “RETRIEVE-FILE-SECURELY” with CRATELO

Figure 3 represents CRATELO’s classes and relationships used to model the RETRIEVE-FILE-SECURELY scenario. For limits of visualization, the diagram covers only the most salient notions involved in this cyber operation.

In order to retrieve a file without exposing a computer system – and possibly an entire network – to cyber threats, some specific security requirements need to be fulfilled while carrying out that operation. In particular, as it is also the case for other kinds of CYBER-OPERATION, RETRIEVE-FILE-SECURELY must occur over a secure channel of a network, from authenticated computer(s) and through authorized server(s). By and large, abiding to these security requirements while executing the mission-tasks should lead to mission accomplishment. The composite RETRIEVE-FILE-SECURELY-TASK can be further divided into simpler temporally-structured and logically-connected subtasks. Accordingly, a request for a file can be sent to an authenticated server only after locating the desired file in the network; the inspection of the file can trivially occur only once the file has been obtained; and so on and so forth. In CRATELO we can express these basic temporal constraints by means of the foundational layer: in fact, DOLCE includes an adaptation of Allen’s axioms [22], which are considered as a powerful logical theory for temporal representation and reasoning (the formalization of these axioms has also been maintained in DOLCE-SPRAY). Moreover, if malware is detected, the file has to be removed from the host: the deployment of this preventive countermeasure aims at avoiding a disruption of the isolated computer node and a cyber attack to the network it belongs to. This countermeasure can be expressed as a conditional rule formalized in CRATELO by using an additional modeling apparatus, i.e. the Semantic Web Rule Language (SWRL) [19], which extends OWL-DL axioms. By including rule-based mechanisms in CRATELO we also comply with the core requisites described in [11] of a full-fledged cyber ontology architecture.

As the example reveals, one of the key design principles underlying CRATELO is to separate the temporal dynamics of cyber operations from the abstract generalizations used to describe them, i.e., plans, tasks, requirements. This approach consents to model a cyber operation as an ‘ontology pattern’ grounded on the top level dyad OCCURRENCE-CHARACTERIZATION and, most importantly, unfolded by the middle-level tetrad [CYBER-OPERATION]-MISSION-PLAN-[SECURITY-REQUIREMENT]. In recent years, ‘ontology patterns’ have become an important instrument in the area of conceptual modeling and ontology engineering [23]; the rationale, as our work suggests, is to identify some minimal knowledge structures to be used as the building blocks for modeling a problem or a scenario in an ontology. This methodology is also ideal from a reasoning point of view. For instance, in [24] the authors state that “mission activities are tasks focused on answering mission questions” (where the latters can be seen as partially overlapping the notion of security requirements): but an ontology that fails to discriminate “activities” from “tasks” would likely be affected in its inference capabilities, in the degree that reasoning over tasks that have not been executed yet – i.e. that are not activities – would not be supported. It’s not difficult to imagine the circumstances where this limit can become a serious drawback for the situation awareness of a cyber analyst: mental simulation is commonly adopted by humans to foresee the outcomes of an action before performing it [25], and a semantic model where mission activities and tasks are conceptually viewed as the same entity precludes that, and might eventually result into pervasive logical inconsistencies, if the ambiguity is not somehow constrained. On the contrary, applying CRATELO’s ontology-pattern allows us to specify cyber operations at a suitable level of conceptual granularity.

[18] Dr. Micheal Susong is an Intelligence Subject Matter Expert affiliated to iSIGHT Partners; he gave an invited talk at Carnegie Mellon University on September 8th, 2014.

IV. CONCLUSIONS AND FUTURE WORK

Notwithstanding the proliferation of taxonomies, dictionaries, glossaries, and terminologies of the cyber landscape, building a comprehensive model of this domain remains a major objective for the community of reference, that includes government agencies, private organizations, researchers and intelligence professionals. There are multiple reasons behind the discrepancy between demand and supply of semantic models of cyber security. Although we can’t thoroughly address this topic here, we are firmly convinced that a great part of the problem is the lack of balance between the ‘vertical’ and the ‘horizontal’ directions of the effort. From one side, state of the art consists of several classifications of the domain, as argued in section II: these efforts typically yields rich catalogs of cyber attacks, exploits and vulnerabilities. On the other side, a rigorous conceptual analysis of the entities and relationships that are encompassed by different cyber scenarios would also be needed, but few work has been done on this horizontal dimension (if we exclude the ongoing MITRE initiative described by Leo Obrst and colleagues in [11]). In this paper we placed ourselves on the second perspective: instead of presenting “yet another” catalog of cyber notions, an endeavor that remains however of undisputable relevance, we decided to explore in depth the semantic space of cyber operations. Our investigation addresses cyber operations as complex entities where the human factor is as important as the technological spectrum: our ontological analysis is grounded on a bedrock of foundational concepts and reaches the domain of cyber operations through an intermediate layer where core notions are defined.

Future work will focus on the following research steps:

- to focus on the ontological analysis of risk;

- to populate the domain ontology with a large set of cyber operations documented in the literature and learned from real-world case studies;
- to design and customize a methodology for ontology validation based on “competency questions” submitted to domain experts (along to what has been proposed in [19]);
- to run cyber warfare simulations within military exercises, collecting data to be modeled with CRATELO;
- to study ontology mappings between CRATELO and other semantic models (e.g. MITRE’s Cyber Ontology Architecture), ensuring interoperability and reusability of the resource.

We are aware of the challenges ahead of us in pursuing this research agenda, which would usually be very difficult to implement. Nevertheless, we’re also persuaded that, in the broad vision framed by the ARL Cyber Security Collaborative Research Alliance, what we have described here illustrates a realistic work plan and a necessary step toward the foundation of a science of cyber security.

ACKNOWLEDGMENTS

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.