Chapter 7
Cybersecurity Ontology

7.1 Introduction

The rapid growth in data through today's digital technologies expands the importance of cybersecurity with regard to the increase of cybersecurity threats, because data are the most important value in the digital world. However, public and private organizations are currently coping with cybersecurity issues without collaboration due to lacks of global standards to solve this problem. Albeit some public and private organizations possess some forms of standards trying to solve this problem based on these standards, which cannot be deployed to fully collaborate with each other. This requires developing ontologies for cybersecurity issues which provides a common understanding of cybersecurity domains. The term ontology itself comes from the Greek words onto, which means existence or being real, and logia, which means science, or study.

Hence, the term ontology specifies some sort of shared understanding. In a more formal sense ontology can be assumed representing some kind of description logic. Furthermore, ontology may indicate that certain object types are subsets of another, and also indicate what can be said about the objects in the respective domains. As an outcome, the ontology can specify which properties each object has, and what value or range of values each property can take. In this regard ontology defines the discourse about that object. Against this background, ontology is a description of what exists specifically in a specific domain, for instance, every component that exists in an information system. This includes the relationship and hierarchy between these components. In this regard the ontology focus is not primarily discussing whether these components are the true essence or core of the information system or not. Furthermore, it is important to note that ontology does not describe whether the components within the information system are more real compared to the process that takes place within the information system. Rather, they are naming components and processes and grouping similar

D. P. F. Möller, *Cybersecurity in Digital Transformation*, SpringerBriefs on
Cyber Security Systems and Networks, https://doi.org/10.1007/978-3-030-60570-4_7

ones together into categories. The purpose of ontology is to understand and describe underlying structures that affect the domain specific components or systems. In this context ontology of a domain specifies the domain specific object, concepts and relations in that domain, which can be assumed as a generally structured description of items. Hence, ontology may also indicate that certain object types are subtypes of another, and specify, which properties each object has, and what value or range of values each property can take. Therefore, ontology of a domain defines the discourse about the domain, and if an item does not appear in ontology, then about that item no statement can be given. In this context, ontology specifies some sort of shared understanding of a domain. In other words the term ontology can be assumed analogous to description logic. Some of the major characteristics of ontologies are that they ensure a common understanding of information and that they make explicit domain assumptions. As a result, the interconnectedness and interoperability of the model make it invaluable for addressing the challenges of accessing and querying data.

7.2 Ontology Types

Ontology is a formal, explicit specification of a shared conceptualization in which the knowledge model can be built upon the following types:

- *Entity*: Represents an object or thing, for example: person, smartphone manufacturer, smartphone user, and many others.
- *Relation*: Represents the relationships between entities, for example: a smartphone manufacturer and smartphone user customer relationship.
- *Role*: Describes the participation of entities in a relation, for example: in a business deal there are roles of manufacturer and user, respectively.
- *Resource*: Represents the properties associated with an *entity* or a *relation*, for example: a name or date, and others. Resources consist of primitive types and values, such as strings or integers.

Against this background ontology specifies the objects, concepts, and relations within the respective domain, and hence can be stated as a structured list of items. In this regard it is a formal naming and definition of types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain of discourse. Moreover, ontology may indicate that certain object types are subtypes of another. Hence, ontology of a domain defines the discourse about that domain. However, if an item does not appear in ontology, then that item cannot be reasoned about. In this regard ontology of a domain specifies one important type of knowledge, for instance, knowledge of the static data in the domain. This includes a vocabulary of terms, definitions of these terms, and a specification of the terms and concepts interrelations. To this extent, ontology specifies some sort of shared understanding of a domain [1]. Hence, ontologies are defined for particular purposes and in particular contents, and the form ontology takes will be at least partially influenced by those purposes and contexts [2]. Moreover, understanding of appropriate domain ontology is a great aid to knowledge acquisition. Thus, ontologies have been designed with different levels of specificity [3].

In recent years, there has been a need to use ontologies in cybersecurity for helping to solve the cybersecurity problem. In [4], the use of Semantic Web Languages and Ontologies (SWLO) for cybersecurity awareness is discussed. Hence, ontologies for cybersecurity go back to the early days of Semantic Web. For instance, in [5] the use of DARPA Agent Markup Language (DAML), the precursor of the Web Ontology Language (OWL) for representing ontology for intrusion detection issues, is discussed. It compares DAML against XML and discusses the inadequacy of the latter. The ontology includes 23 classes and 190 properties/attributes. OWL is a semantic web computational logic-based language, designed to represent rich and complex knowledge about things and the relations between them. It also provides detailed, consistent and meaningful distinctions between classes, properties and relationships. By specifying both, object classes and relationship properties as well as their hierarchical order, OWL enriches ontology modeling in semantic graph databases, also known as Resource Description Framework (RDF). RDF is a model for data publishing and interchange on the Web standardized by the World Wide Web Consortium (W3C). In this regard, RDF triplestore is a type of graph database that stores semantic facts. OWL, used together with the OWL reasoner in RDF triplestores, enables consistency checks to find any logical inconsistencies, and ensures satisfiability checks to find whether there are classes that cannot have instances. The data in a RDF triplestore is stored in three linked data pieces which are called a triple. Triples are also referred to a statement or RDF statements [6]. Also, OWL is equipped with means for defining equivalence and difference between instances, classes and properties. These relationships help users match concepts even if various data sources describe these concepts somewhat differently. They also ensure the disambiguation between different instances that share the same names or descriptions [6].

7.3 Cybersecurity Ontology

The rapid growth in data through today's digital technologies expands the importance of cybersecurity with regard to the increase of cybersecurity threats, because data is the most important value in the digital world. In this context, this data is available in structured, semi-structured, and unstructured forms for both, data from internal and external sources. Therefore, unifying such scattered data will provide better visibility and situational awareness with regard to cybersecurity analysis as well as a more proactive and possibly predictive approach to avoid cyber threats. Against this background the development of cybersecurity attack ontology is essential to enable the secure data integration across disparate data sources. In this context cybersecurity attack ontology is required for modeling different types of adversary knowledge. Therefore, security attack ontology aims at building a knowledge base for security attacks that describe type, mode, consequences, and others. Developing the cybersecurity attack ontology one can make use of known security standards (see Sect. 6.3), for instance ISO/IEC 15408:2009, ISO/IEC 18045, ISO/IEC 27000:2012, ISO/IEC 17799:2005, NIST SP-800:30, and others, and security dictionaries (see Sect. 6.3), for instance Common Vulnerabilities and Exposures (CVE), CAPEC™, OWASP, Comprehensive Lightweight Application Security Process (CLASP), and others. Thus, the security ontology can make use of the foregoing constructs with regard to the Web Ontology Language (OWL), a language for defining ontologies to describe properties of web resources [7]. OWL is a semantic web computational logic-based language, designed to represent rich and complex knowledge about things and the relations between them. It also provides detailed, consistent and meaningful distinctions between classes, properties and relationships. OWL based ontology describes a domain in terms of classes, instances and relations and include descriptions of the characteristics of those objects with regard to slots and internal links such as instance-of and subclass-of. Based on the conceptual aspects about attack models and attack scenarios presented in Chap. 6 and the security standards, the cyber security attack ontology can be illustrated as shown in Fig. 7.1.

In the context of semantics it is possible to execute precise searches and complex queries. Initially, this effort is focused on cyber threat malware subjects, because malware is one of the most prevalent cyber threats to cybersecurity. For this reason the MITRE Corporation has developed the Malware Attribute Enumeration and Characterization (MAEC) language [8] which is a structured language for encoding and sharing high-fidelity information about malware based upon attributes such as behaviors, artifacts, and relationships between malware subjects. As described in [8], MAEC focuses on characterizing the most common malware types, including Trojans, worms, rootkits, and many others, as well as today's more advanced malware types. MAEC's core components include a vocabulary, a grammar, and a standardized output format, and provide a standard means of communicating information about malware attributes, as shown in Fig. 7.2.

Before MAEC, the lack of an accepted standard for unambiguously characterizing malware subjects meant there was no clear method for communicating the specific malware attributes detected in malware by the analyses, or for enumerating its fundamental makeup. The results included non-interoperable and disparate malware reporting between public and private organizations, disjointed or inaccurate malware attribution, the duplication of malware analysis efforts, increased difficulty in determining the severity of a malware threat, and a greater delay between malware infection and detection as well as response [9]. However, the key to ontology development is an understanding of the cyber domain, which drives the kinds of entities, properties, relationships, and potentially rules that will be needed in the ontology. With regard to the complexity of cybersecurity analysis, the ontology development better consist of modular sub-ontologies, rather than a single, monolithic ontology [10]. Thus ontologies can be grouped categories such as upper level ontologies, mid-level ontologies, and domain level ontologies, according to their specific levels of abstraction of the respective cybersecurity architecture ontology concept to be developed. For more details see [10–12].

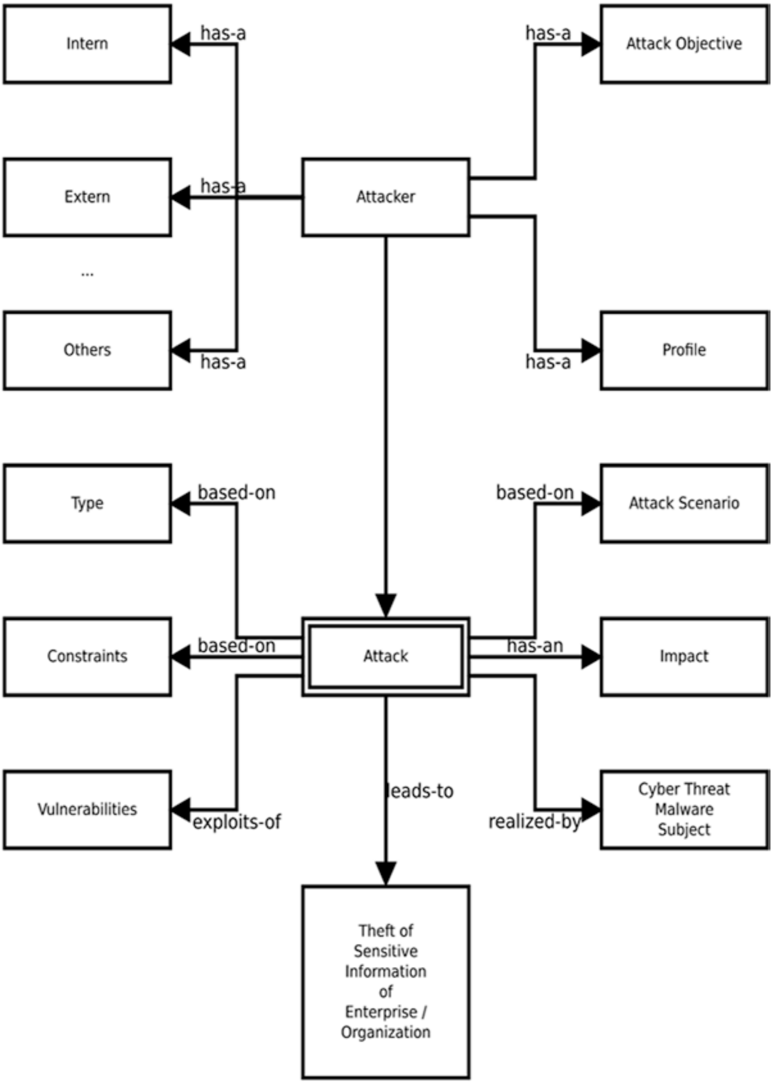


Fig. 7.1 Cyber security attack ontology

Developing the detailed architecture of the cybersecurity ontology requires, dependent of the category of interest, specific descriptions to abstract major categories, domain specific concepts, and ontologies that span multiple concept categories. The descriptions of the major categories which lay the basis for a cybersecurity ontology taxonomy are:

- *Entities*: Describe foundational incidents, collections, and others.

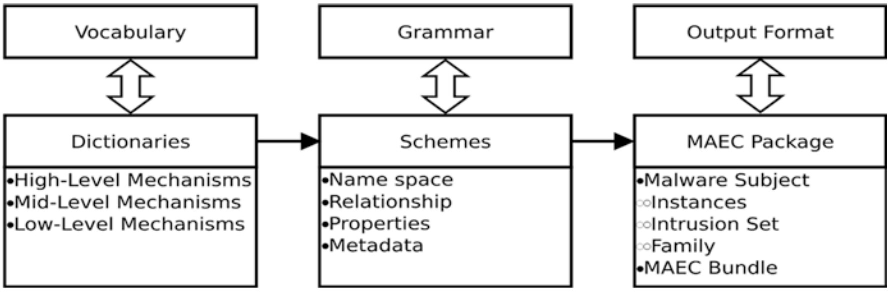


Fig. 7.2 MAEC's core components vocabulary, grammar, and output format

- *Relations*: Describe relationships of detection and defense actions, organizational locations, and others.
- *Role*: Describe cyber threat attackers and cyber threat defenders.
- *Resources*: Describe capability, infrastructure, behavior, malware subjects, and others.

In regard to malware, resources published that attempt to systematically categorize malware subject's ontology are reported in [13], and descriptive languages implemented in Extensible Markup Language (XML) in [7, 14, 15]. The ontology described enables data exchange between security algorithms. Their taxonomy of malware classes is shown in Fig. 7.3. Also, worthy to mention is an attempt of categorizing malware subject traits [16]. This development finally ended up in the so-called Unified Cybersecurity Ontology (UCO) framework described in [16] helping to evolve cybersecurity standards from a syntactic representation to a more semantic representation showing several contributions for the cybersecurity ontology. In this regard UCO is an extension to Intrusion Detection System ontology [5] to describe incidents related to cybersecurity. Several projects that focus on individual components of a Unified Cybersecurity Ontology framework analyze different data streams and assert facts in a so called triplestore approach, as reported in [5, 17, 18]. In this context UCO is essential for unifying information from heterogeneous sources and supporting reasoning and rule writing. Thus, UCO supports reasoning and inferring new information from existing information, and also supports capturing specialized knowledge of cybersecurity analysts which can be expressed using ontology classes and terms as well as rules.

Besides OWL language, the MITRE Corporation has launched the Malware Attribute Enumeration and Characterization (MAEC) language [7], a structured language for encoding and sharing high-fidelity information about malware subjects based upon attributes such as behaviors, artifacts, and relationships between malware subjects. Malware is responsible for a variety of malicious activities, ranging from spam email distribution via botnets to the theft of sensitive information via targeted cyber threat attacks. Therefore, the protection of computer systems and networks from malware is a primary cybersecurity concern for public and private organizations, as even a single instance of uncaught malware can result in damaged

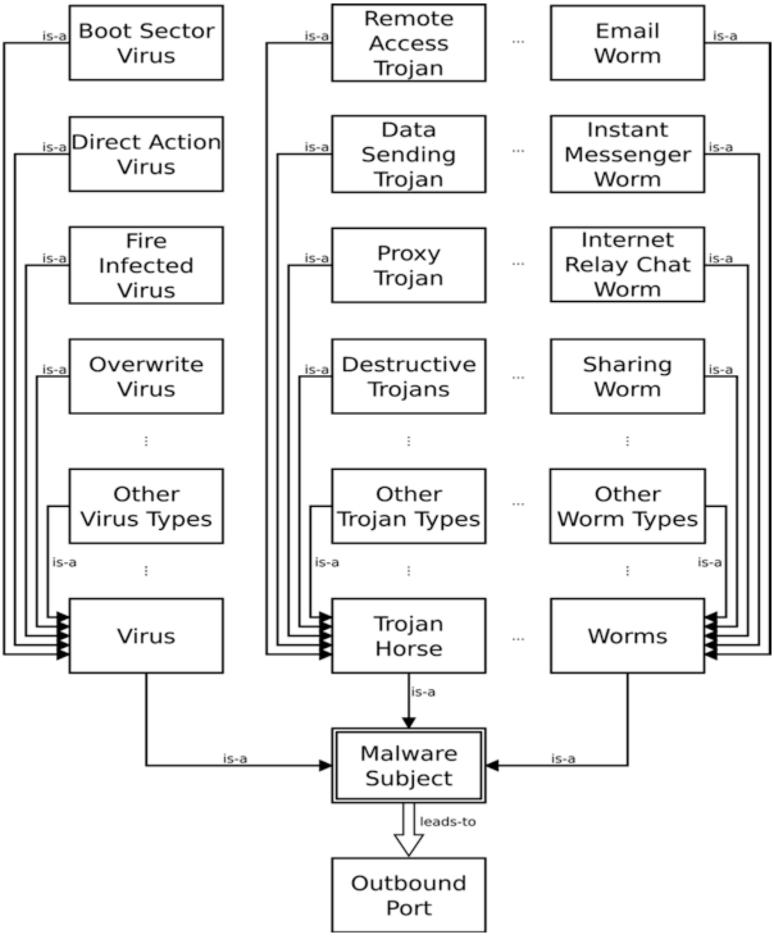


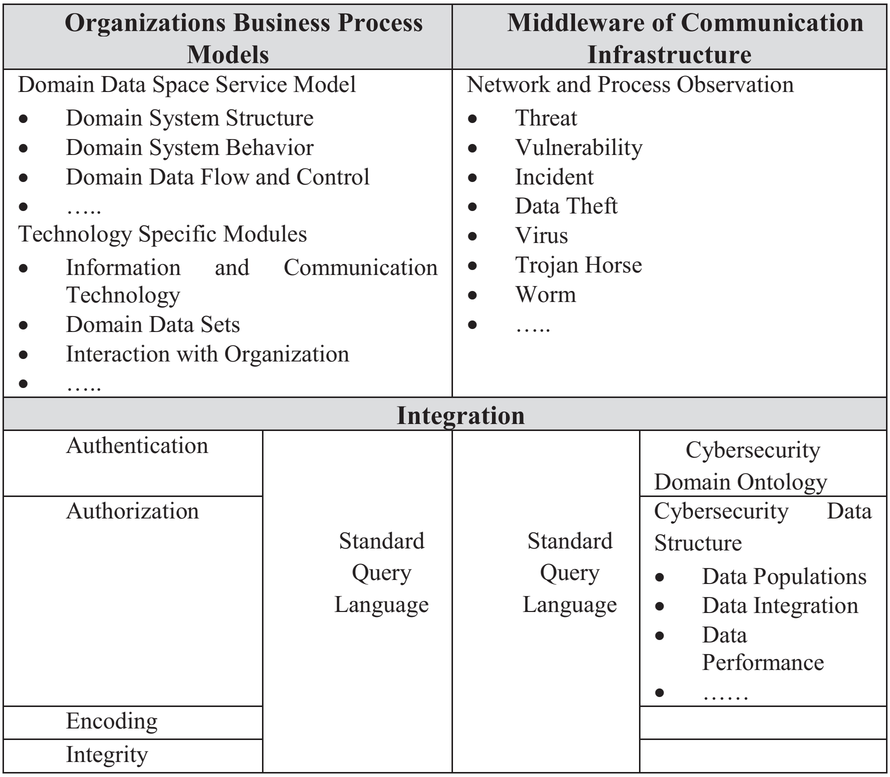
Fig. 7.3 Taxonomy of malware classes

computer systems and compromised data. However, the key to ontology development is an understanding of the respective cyber domain, which drives the kinds of entities, properties, relationships, and potential rules essential to the cybersecurity ontology.

Against this background the cybersecurity ontology framework includes the cybersecurity domain-specific ontology and data integration for different data sources in a common knowledge base, for instance, metadata records. This enables data integration and padding from ontology information and access to various data sets. This also has to include security services related to the respective business processes, network devices, and the requirements ultimately required to provide cybersecurity against cyber threat attack incidents as part of the middleware. The integration is the required interaction between data set infrastructure and cybersecurity ontology layers that provide the requirements for cybersecurity to detect cyber threat attack incidents, prevent cyber threat attack incidents, and avert cyber threat attack incidents for developing cybersecurity domain-specific ontology.

With respect to cyber threat attack intrusion incidents on data sets, it is assumed that a cyberattack maps to the category unknown, pointing to unpredictable and unexpected cyber threat attack incidents. This represents a dynamically changing risk for the data space, in the digital transformation era, which requires an adequate solution for unpredictable incidents to make data space cyber-secure. This requires the domain-specific semantics of unknowns as a kind of uncertainty that must be represented by their ontologies. Such ontologies must be able to suggest suitable cybersecurity services that may or may not be required, which have to be set at design time of the data record, and customized and activated by the data sets used. Hence, the architecture of a generic cybersecurity ontology framework is based on components, as shown in the generic model in Fig. 7.4.

The generic cybersecurity ontology framework in Fig. 7.4 shows the essential system components, including the cybersecurity domain-specific ontology and data integration for different data sources in a common knowledge base, for instance



<table>
<tr><th colspan="2">Organizations Business Process Models</th><th colspan="2">Middleware of Communication Infrastructure</th></tr>
<tr><td colspan="2">Domain Data Space Service Model
• Domain System Structure
• Domain System Behavior
• Domain Data Flow and Control
• …..
Technology Specific Modules
• Information and Communication Technology
• Domain Data Sets
• Interaction with Organization
• …..</td><td colspan="2">Network and Process Observation
• Threat
• Vulnerability
• Incident
• Data Theft
• Virus
• Trojan Horse
• Worm
• …..</td></tr>
<tr><th colspan="4">Integration</th></tr>
<tr><td>Authentication</td><td rowspan="4">Standard Query Language</td><td rowspan="4">Standard Query Language</td><td>Cybersecurity Domain Ontology</td></tr>
<tr><td>Authorization</td><td>Cybersecurity Data Structure
• Data Populations
• Data Integration
• Data Performance
• ……</td></tr>
<tr><td>Encoding</td><td></td></tr>
<tr><td>Integrity</td><td></td></tr>
</table>

Fig. 7.4 Generic cybersecurity material data space ontology framework

metadata sets. This enables data integration and padding from ontology information and access to various data sets required. Furthermore, security services related to organizations business process models, network devices, and the requirements ultimately required providing security against cyber threat incidents.

Data sets integration in the middleware layer provides the requirements for cyber security, for instance cyber threat attack intrusion incidents and other vulnerabilities that create the security framework for using domain-specific cybersecurity ontologies.

The queries combining all values of data sets that are analyzed for cyber threat attacks using cybersecurity domain-specific context ontologies. The queries for identifying possible cyber threat attack incidents according to the architecture shown in Fig. 7.4 must, as described, characteristically map in depth the underlying information of the considered data sets in ontologies so that cybersecurity can map domain-specific ontologies to it.

The cybersecurity core ontologies form, in a certain sense, strengths and weaknesses profiles that map the security requirements to the possible entities. The ontology for cyber secure operations aims to reduce potential false positives in detecting potential cyberattacks that may arise when monitoring cyber vulnerabilities. Thus, cybersecurity ontology represents a domain-specific model that defines the essential domain concepts, their properties and the relationships between them and represents an essential knowledge base to cyber secure the respective application. The generic cyber-attack model is shown in Fig. 7.5.

As shown in Fig. 7.5 cyber threat analysis is a security field that needs a more scientific basis for sharing information among cyber defending teams. One option is building OWL-based malware analysis ontology to provide that more scientific approach based on a malware analysis dictionary and taxonomy, and combining those in a competency model with the goal of creating an ontology-based cybersecurity framework. Meanwhile several security standards have been developed, taking into account OWL, representing ontology based security concepts such as: incident reporting, threat information, risk information, assets, target information. Each group contains multiple metrics, also known as *factors,* used to compute a Common Weakness Security System (CWSS™) score for weaknesses. CWSS™ is co-sponsored by the MITRE Corp. [18]. Thus, ontology can be defined as abstract representation of real-world objects, which means that ontology constitutes a domain-specific model defining the essential domain concepts, their properties, and the relationships between them, represented as a knowledge base.

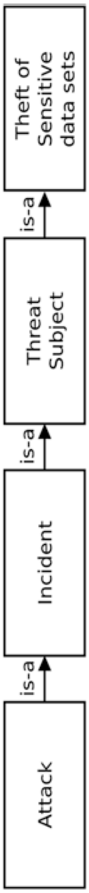


Fig. 7.5 Cyber-attack model

References

1. M. Uschold, Knowledge level modeling: concepts and terminology. Knowl. Eng. Rev. **13**, 5–29 (1998)
2. B. Chandrasekaran, J.R. Josephson, V.R. Benjamins, The ontology of tasks and methods, in *Proceedings of the 11th Banff Knowledge Acquisition for Knowledge for Knowledge-Based System Workshop*, 1998
3. N. Sadbolt, K.O. Hara, H. Cottam, The use of ontologies for knowledge acquisition, in *Knowledge Engineering and Agent Technology*, ed. by J. Cuena, Y. Demazeau, A.G. Serrano, J. Treur, (IOS Press, Amsterdam, 2004), pp. 19–42
4. A. Sheth, Can semantic web techniques empower comprehension and projection in cyber situational awareness, in *ARO Workshop*, 2007
5. J. Undercoffer, J. Pinkston, A. Joshi, T. Finn, A target-centric ontology for intrusion detection, in *18th International Joint Conference on AI*, 2004, pp. 9–15
6. https://www.ontotext.com/knowledgehub/fundamentals/what-are-ontologies/
7. S. Bechhofer, OWL: web ontology language, in *Encyclopedia of Database Systems*, ed. by L. Liu, M.T. Özsu, (Springer Publ., New York, 2009). https://doi.org/10.1007/978-0-387-39940-9_1073
8. MAEC—Malware Attribute Enumeration and Characterization. http://maec.mitre.org/
9. http://maecproject.github.io/about-maec/
10. L. Obrst, P. Chase, R. Markeloff, Developing an Ontology of the Cyber Security Domain. http://ceur-ws.org/Vol-966/STIDS2012_T06_ObrstEtAl_CyberOntology.pdf
11. L. Obrst, Ontolological architectures in Theory and Applications of Ontology - Computer Applications, ed. by J. Seibt, A. Kameas, R. Poli, Chapter 2, pp. 27–66, (Springer Publ. London, 2010)
12. S. Semy, M. Pulvermacher, L. Obrst, Toward the Use of an Upper Ontology for U.S. Government and U.S. Military Domains: An Evaluation, MITRE Technical Report, MTR 04B0000063, 2005
13. M. Swimmer, Towards an Ontology of Malware Classes. http://www.scribd.com/doc/24058261/Towards-an-Ontology-of-Malware-Classes
14. IEEE-SA—Industry Connections. http://standards.ieee.org/develop/indconn/icsg/malware.html
15. MANDIANT: Intelligent Information Security. http://www.mandiant.com
16. L. Zeltser, Categories of common malware traits, Internet Storm Center Handler's Diary, 2009. http://isc.sans.edu/diary.html?storyid=7186
17. S. More, M. Matthews, A. Joshi, T. Finn, A knowledge-based approach to intrusion detection modeling, in *Proceedings of the IEEE Symposium on Security and Privacy Workshops*, 2012, pp. 75–81
18. https://cwe.mitre.org/cwss/cwss_v1.0.1.htm