Towards Ontology-based Cyber Threat Response

Nikolay Kalinin

Faculty of Computational Mathematics and Cybernetics,
Lomonosov Moscow State University,
119991, GSP-1, 1-52, Leninskiye Gory, Moscow, Russia

Abstract. Response to the threats of information security in conditions of modern organization with a large infrastructure is an area with emergency loaded intensity of the data usage. For a successful exposure and the prevention of computer attacks the construction of complex models of the events is required. In this work, the question of the applicability of ontological models is examined for the description of threats. On the basis of worked out applied ontologies, the model architecture of the knowledge base is being offered, the possible practical scenarios of its use are being examined. The peculiarities of this work are the usage of reasoning on the different stages of event handling and design of knowledge, not only about events but also about an information infrastructure and its safety. Thus, the examined semantic technologies can be a base for the complete system of response to the threats of information security.

Keywords: Ontology · Reasoning · Cyber security · Threat response

1 Introduction

An area of information security today is especially relevant: the amount of threats and their destructive capacity grow with every year. Computer attacks are complicated trigger able operations in that it can involve considerable amount of network nodes now. Intruders use the various techniques of conducting attacks and concealment of their activities, complicating work of defenders at the same. In such circumstances, the development of new methods that would be applied in composition with the automated tools of exposure of cyber threats becomes not simply an interesting scientific task, but also a valuable practical result.

The traditional approach for the exposure of threats is based on the use of signatures, namely search of suspicious templates in operating data. The signature approach is ubiquitously used due to the ease and profitability, from the point of view of computing resources. Unfortunately, it has a range of substantial drawbacks: the signature approach requires considerable efforts on maintenance of base of signatures in the actual state not being able to expose new types of threats (zero-day attacks) and does not allow to line up the models of complex attacks.

The most popular alternative for the signature approach is the usage of methods of machine learning. A search of threats, which is based on the exposure of

anomalies, behavioral, and statistical analysis, allows us to find out substantially more difficult attacks than the signature approach, but it is not deprived of the defects. Machine learning algorithms results are often poorly interpretable thus there are difficulties in localization and threat removing. In addition, such algorithms often require fine-tuning under a correct infrastructure and skilled support for the timely account of inevitable changes of external terms.

To two indicated approaches we can add and approach on the basis of formal models. As noted in the work [12] ontology is already one of the fixed assets of realization of the large systems of information security because it allows us to use the experience of wide expert association for providing of transparency of work and forecast of results. Tools of exposure threat on the basis of formal models can allow not only to identify and classify threats but also to effectively produce reliable and interpreted decisions for their removal. The key advantage of such tools is a higher level of the used abstractions that provides knowledge systematization, decision automation, and allows to offer to the expert the decisions with the observance of formal response procedures.

One of the problems of conceptual models usage in dynamically developing areas is the laboriousness of knowledge base maintenance in the actual state. But the wide usage of open and community-supported standards and taxonomies, such as CVE/NVD [1] or CAPEC [2] allows to avoid the difficulties related to updating of the knowledge base. On the one side, open and peer-reviewed sources allow where appropriate to specify the terminological base of ontology and on the other, allow to update the actual filling of the knowledge base regularly. Another unique advantage of using formal models in cybersecurity is the high development level of industrial systems for collecting information about protected objects. In modern organization all required information is already presented in inventory databases and SIEM systems [3], that considerably simplifies its integration in the knowledge base.

In spite of the fact that ontology a long ago and successfully used in many areas, such as genetics and bio-medicine, they meet rarely in enterprise solutions providing information security. The purpose of this work is a demonstration of wide possibilities of an ontological approach for the development of methods and tools for reacting to the security threats of the distributed information infrastructure. Central directions of our research are the questions related to applicability and efficiency of logical reasoning and also questions related to the conceptual representation of knowledge about an information infrastructure.

The brief review of accessible works is given in the second part of this article, the third part contains a scheme description of the model knowledge base, on which in fourth part some possibilities of ontological approach are demonstrated.

[1] https://cve.mitre.org
[2] http://capec.mitre.org
[3] SIEM (Security information and event management) - class of the systems carrying out the centralized collection and analysis of security log

2 Related Works

The construction of ontological models in information security is conducted already for more than fifteen years. One of the first bright works in this area is ontology IDS [4], presented in [17]. Authors put the aim to show the utility of ontology as a model for classification of attacks in the intrusion detection system underlining their superiority above more used taxonomies due to greater flexibility and the possibility to work with heterogeneous data. Their result ontology presented as attack classification framework and described in DAML-OIL [3] (language predecessor OWL [7]) ontology plugging in more than 190 concepts and operating with the data got as a result of instrumentation of the Linux kernel. Note that one of the dignities of the built ontology authors count the unambiguity of objects distribution on classes, thus, the same high level of strictness is arrived at, as well as at reasoning based on the use of taxonomies. Possibilities of the use of the built model are on example SYN flood attacks and buffer overflow. The classification consists of a selection of the most correct class that would correspond to the happening event.

Another classic example of the ontological model usage is presented in the article [6], in that it is suggested to use the ontology of information security for annotating functional features web of resources. Final ontology appears as seven sub ontologies and intended for description of security mechanisms such as protocols, algorithms, and registration data. In a difference from IDS of ontology that is intended for the use in a certain application, the ontology of submitted authors is a general ontology of information security and can be used for annotating any the web of resources.

Development of semantic models is directly connected with the use of industry standards and specifications, so in work [18] OVM ontology based on taxonomies is presented and standards of corporation MITRE [5] (CVE, CWE, CAPEC, CVSS) and intended for description of weakness in software products. The built ontology is one of maiden attempts to bind the current standards of description in a more difficult and complete model.

Other example of the successful use of open dictionaries is described in [5]. By a basic problem at the automated use of such data authors consider a presence of important information presented as text. The result of their research is framework trained for extraction of relevant content. Extracted entities interconnection between them appears as RDF-triplets on the basis of simple ontology, complementary of IDS [17] ontology. The final system is integrated into the infrastructure of the linked open data (LOD) [6] .

Approach allowing to systematize not only information security but also the development of ontology process, presented in works [8] and [9]. In the first authors examine methodology of construction of ontology in cybersecurity. The construction of ontology in their opinion consists of the next stages:

[4] IDS - Intrusion detection system
[5] https://mitre.org
[6] https://lod-cloud.net/

1. Determination of the aims shown in the required queries to the knowledge base and supposed scenarios of the use.
2. Analysis of existent ontologies of the same subject domain including all valuable concepts from them here. If the number of concepts is great authors recommend to include whole ontology in the complement of the developed scheme.
3. Addition of connections coming from data with that it is assumed to work and coming from necessities and existent industry standard.

In authors' opinion, ontologies are usually an association of three levels, from most general, such as DOLCE, at the top level, to the applied ontology [9], this approach gets further development described as full ontology-framework CARTELO. The ontology DOLCE- SPRAY is used at the top level, at middle-level ontology is presented by the ontology of SECCO, plugging in itself the basic concepts of cybersecurity, the ontology of cyber-operations OSCO complements at the bottom level.

By the natural desire of researchers, that in the total got the embodiment in a number of works, was to overcome one ontology of all traditional scenarios of the use of concepts of cybersecurity. Thus in work [2] there is an example of the complex use of ontology made in composition the system of cybersecurity. The Package-oriented ontology for the description of network traffic of PACO is used as a kernel for extraction of knowledge from network traffic and as an instrument of classification of traffic and, together with more general top-level ontologies (CARTELO), as an interface for analyst work. In stand experiments where efficiency of the system was compared for the exposure of attacks with the and without use of ontology advantages of ontological approach were shown. In the total authors come to the conclusion that combination of high-level ontologies and low-level ontologies allows to substantially increase expressiveness of semantic model, and usage of such models together with traditional tools to become the basis for the system of decision making, the superior possibility of analyst.

In works [16] and [15] an example is made not simply constructions of ontology, but also developments of the architecture of knowledge base for its use. As basic functional components of the system authors distinguish the component of incidents handling presented by the bases of incidents and warnings; asset management component, presented by the base of resources; and the component of accumulation of knowledge. The last includes the knowledge base of products and services, the knowledge base of risks, and the knowledge base of countermeasures that contain knowledge based on the treatment of industry standards. Ontology, here, is a tool for uniform manipulation of the collected heterogeneous data.

In [4] authors note that formal representation of knowledge and integration of information from different sources allows substantially improve quality of exposure and response. In the article as main scenarios of the use are the search of relevant records from IDS, collection of information about software, and attempt of determination of malicious activity on the basis of network traffic and

changes in the system. For the solution of these tasks, authors develop ontology of STUCO. Its notable features are relative simplicity and realization by means of JSON- scheme from one side, promotes its practical applicability, but with other lays on substantial limitations, main from that is the impossibility of the logical reasoning mechanism usage

The common decision of long-term problem standardization of formats of cybersecurity-related knowledge lately became language STIX [1], therefore no wonder that the most complex is universal ontology of cybersecurity (UCO), presented in work [14] is based on exactly its structure. An offered ontology is implemented in OWL DL assuming an effective inference allows to extract information from all popular industrial dictionaries and assumes the wide spectrum of scenarios of the use. Meantime its valuable use in practice feasible only after its adaptation to certain tasks by means of the addition of corresponding applied ontologies. UCO is the most successful attempt to create a middle-level ontology, that from one side would possess sufficient expressiveness for the description of conceptions of any cybersecurity directions and with other abandoned space for the clarification of bottom level ontologies.

In the conclusion of this review, we want to note that in spite of the fact that for the past years substantial results were obtained with the area of development of cybersecurity ontologies many tasks are not solved. Possibility of reasoning is not used even in those works where implementation allows to use them. The problem of extraction of knowledge from the unstructured sources is not fully resolved although work makes considerable part of analysing such data. Ontologies do not contained concepts for description of information infrastructure in the meantime the question of cybersecurity prioritization events is continuously related to such knowledge. A possible way for efficient infrastructure representation presented in recent work [10], but valuable ontologies containing both knowledge about infrastructure and knowledge of information security are yet to be developed.

3 Knowledge Base Architecture

To show the possibilities of ontological approach the model knowledge base was implemented. A terminological constituent (T - box) of knowledge base is ontology of UCO complemented applied with bottom level ontologies. An actual constituent (A - Box) plugs operating information (events and incidents of cybersecurity), information about an infrastructure and also information from open dictionaries and taxonomies.

3.1 Ontological Model

As said earlier UCO though and is the most complete cybersecurity ontology in pure form fits badly for practical application and requires adaptation. As such adaptation, additional ontologies for the decision of certain tasks were developed. Ontology of operating information extends and complements such concepts of

UCO as action and observable. Its main task is to provide accordance with other objects of knowledge base and by operating information. Its key concept is the event. The event is the universal observed object and parent for all other types, which represent events in the real world. In addition, it plugs in description rules of threats exposure (signatures, anomalies, and others) and sets their accordance with industry standards, such as a matrix of ATT&CK[13]. Ontology of information infrastructure is a clarification for uco – identity - local identity that allows to determine authentication for internal subjects and also essence set of infrastructure objects for description of endpoints and applications in an infrastructure (frequently by the subclasses of uco - observable) and their well-known and possible connections. Last from ontology models is prioritization ontology. It is a model for a conclusion of environmental risk metrics CVSS. It includes concepts from the environmental risk of CVSS. Requirements to confidentiality, availability, integrity, probability are causing damages that hatch on the basis of data about subject to the risk to the infrastructure.

3.2 Presentation of Operational Information

In a model knowledge base operating information appears as events of SIEM because the systems of this class are the main components of the centralized security monitoring in enterprise surroundings. From the point of data view, SIEM events are the records of compatible format, extracted from different logs and security tools aggregated in a single database. The format of records is based on the Cybox standard (http://cybox.mitre.org), plugged into STIX. As a model data, the logs of regular subsystem of audit of OS Linux, logs generated by Osquery framework (https://osquery.io), and logs of firewalls were used.

3.3 Infrastructure Presentation

Information about an infrastructure appears in two basic types of objects: endpoint record and network rule. The first type contains information about a certain host, such as the installed software, security policies, criticality of the processed information etc. The second is an object for network availability description and written down like the rules of firewalls (that make the basic filling of this part of database) with the only exception that except the standard types of Deny and Allow the type of Routine is intended for description beforehand of well-known permanent network connections.

4 Use Cases

4.1 Attack Classification

We will consider the mechanism of attack classification with the example of event finding out reverse shell on the host, detected by the system of traffic analysis. Initially an event is a record of SIEM and rule of sensor associated

with it. The task of classification, in this case, can be reformulated in terms of conceptual model as a task of search of the most certain concept for this event of SIEM would satisfy description of that. The tree of specification of class for our example is brought around to Fig. 1.

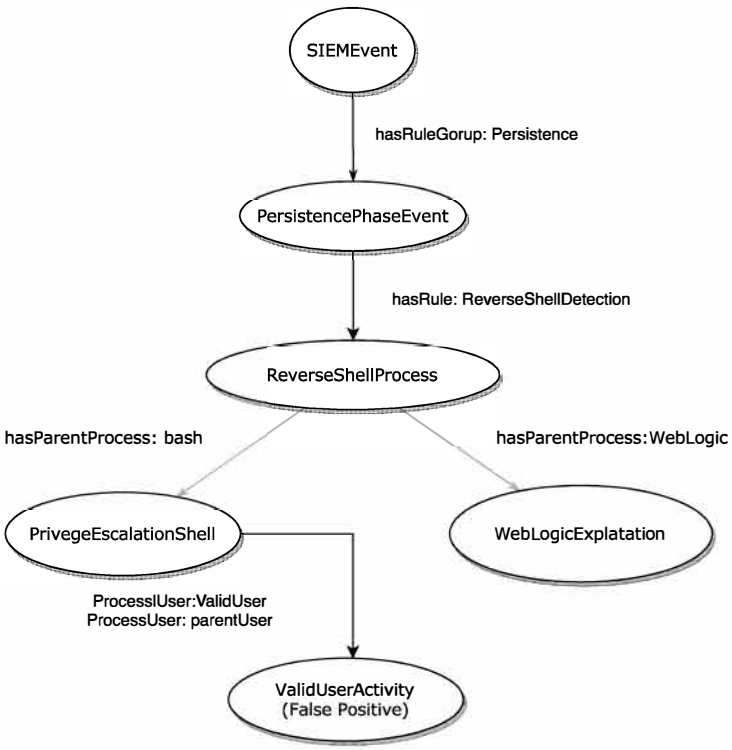

Fig. 1. Logical scheme of class construction for an event

Thus, classification on the basis of reasoning can be basis not only for a decision-making by a man, but also for the acceptance of the automated decision. In our example, such solution is automatic filtration of false positive.

4.2 Risk Assessment

In our model, a risk level is estimated in accordance with the second version CVSS standard. The standard of CVSS is plugged in itself by three types of metrics: base, temporal, and environmental. The first two metrics are descriptions of vulnerability presented in ontology as the property hasCVSSScore and can be delivered from the open-source. The third metrics group is intended for bringing

resulting amendments taking into account descriptions of the information environment and their calculation makes the most interest. For the calculation of environmental metrics descriptions of the affected objects are used. So relation belongToSystem of class Ednpoint allows defining requirements for confidentiality availability and integrity, coming from properties of the system such as a type of processed information and degree of criticism. Probability of indirect damage settles accounts coming from criticism of the constrained systems and closeness of aims on the basis of amount of hosts on that the vulnerable version

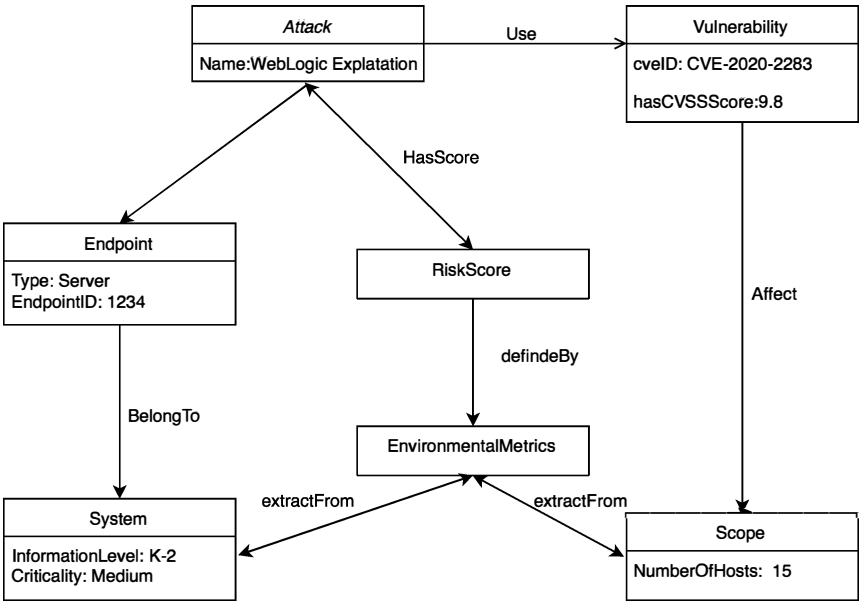

Fig. 2. Logical scheme of risk assessment

4.3 Finding of Related Information

Finding of related information in our model can be materialized on the basis of rules presenting as SPARQL [11] queries. So, for example, for the event of finding out a ssh-tunnel the important constrained information is: information about a source, information about a purpose, information about prohibitive or permitting such connection rules. Such SPARQL queries must be certain for every type of event at the level of the user interface. It is needed to notice that

information that is required for the automated treatment does not fall into a category constrained and hatches by means of mechanisms of ontology.

5 Conclusions and Directions for Further Work

Within the work the model of knowledge base was built to support processes of response to the threats of information security. Stand tests on the basis of model scenarios of the use showed possibility of deployment of ontological approach in the process of response to the incidents of information security. The special attention at development of ontological model was spared to description of information infrastructure as modern processes of providing information security in large organizations indissolubly connected with the processes of network control and eventual devices. Despite the fact that the ontological model has shown its suitability, there is still a long way to go for its full use. Firstly, in work we did not involve the question of possibility of thread data processing and, as a result, the productivity questions, including questions that are related to the choice of optimal dialect of OWL for description of model. Secondly, a fairly primitive model for describing network availability was used, in that the question of presence or incommunication, in fact, is taken to the presence of corresponding rule on the firewall was used. Thirdly, valuable use of the system is impossible without serious expansion of types of processed events and expansion of set of concepts in ontologies of application layer. Our global aim is to develop complete ontological framework for support of response to cyberthreats and this research is only the first step on a path to this aim.

Aknowledgements. This work is supervised by Nikolay Skvortsov, Federal Research Center Computer Science and Control of the Russian Academy of Sciences (FRC CSC RAS).

References

1. Barnum, S.: Standardizing cyber threat intelligence information with the structured threat information expression (stix). Mitre Corporation **11**, 1–22 (2012)
2. Ben-Asher, N., Oltramari, A., Erbacher, R.F., Gonzalez, C.: Ontology-based adaptive systems of cyber defense. In: STIDS. pp. 34–41 (2015)
3. Horrocks, I., et al.: Daml+oil: A description logic for the semantic web. IEEE Data Eng. Bull. **25**(1), 4–9 (2002)
4. Iannacone, M., Bohn, S., Nakamura, G., Gerth, J., Huffer, K., Bridges, R., Ferragut, E., Goodall, J.: Developing an ontology for cyber security knowledge graphs. In: Proceedings of the 10th Annual Cyber and Information Security Research Conference. pp. 1–4 (2015)
5. Joshi, A., Lal, R., Finin, T., Joshi, A.: Extracting cybersecurity related linked data from text. In: 2013 IEEE Seventh International Conference on Semantic Computing. pp. 252–259. IEEE (2013)
6. Kim, A., Luo, J., Kang, M.: Security ontology for annotating resources. In: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". pp. 1483–1499. Springer (2005)

7. McGuinness, D.L., Van Harmelen, F., et al.: Owl web ontology language overview. W3C recommendation **10**(10), 2004 (2004)
8. Obrst, L., Chase, P., Markeloff, R.: Developing an ontology of the cyber security domain. In: STIDS. pp. 49–56 (2012)
9. Oltramari, A., Cranor, L.F., Walls, R.J., McDaniel, P.D.: Building an ontology of cyber security. In: STIDS. pp. 54–61. Citeseer (2014)
10. Scarpato, N., Cilia, N.D., Romano, M.: Reachability matrix ontology: A cybersecurity ontology. Applied Artificial Intelligence **33**(7), 643–655 (2019)
11. Sirin, E., Parsia, B.: Sparql-dl: Sparql query for owl-dl. In: OWLED. vol. 258 (2007)
12. Sokolov, I., Kupriyanovsky, V., Namiot, D., Sukhomlin, V., Pokusaev, O., Lavrov, A., Volokitin, Y.: Modern eu research projects and the digital security ontology of europe. International Journal of Open Information Technologies **6**(4), 72–79 (2018)
13. Strom, B.E., Applebaum, A., Miller, D.P., Nickels, K.C., Pennington, A.G., Thomas, C.B.: Mitre att&ck: Design and philosophy. Technical report (2018)
14. Syed, Z., Padia, A., Finin, T., Mathews, L., Joshi, A.: Uco: A unified cybersecurity ontology. In: Workshops at the Thirtieth AAAI Conference on Artificial Intelligence (2016)
15. Takahashi, T., Kadobayashi, Y.: Reference ontology for cybersecurity operational information. The Computer Journal **58**(10), 2297–2312 (2015)
16. Takahashi, T., Kadobayashi, Y., Fujiwara, H.: Ontological approach toward cybersecurity in cloud computing. In: Proceedings of the 3rd international conference on Security of information and networks. pp. 100–109 (2010)
17. Undercoffer, J., Joshi, A., Pinkston, J.: Modeling computer attacks: An ontology for intrusion detection. In: International Workshop on Recent Advances in Intrusion Detection. pp. 113–135. Springer (2003)
18. Wang, J.A., Guo, M.: 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. pp. 1–4 (2009)