

Making Sense of Future Cybersecurity Technologies: Using Ontologies for Multidisciplinary Domain Analysis

Claire Vishik¹ · Marcello Balduccini²

¹Intel Corporation
claire.vishik@intel.com

²Drexel University
marcello.balduccini@drexel.edu

Abstract

Security experts have difficulties achieving quick vulnerability mitigation because cybersecurity is a complex multi-disciplinary subject that yields itself with great difficulty to traditional methods of risk analysis. In particular, the effectiveness of mitigation strategies depends on an accurate understanding of the relationships among the components of systems that need to be protected, their functional requirements, and of the trade-off between security protection and core functionality. Mitigation strategies may have undesired ripple-effects, such as unexpectedly modifying functions that other system components rely upon. If some of the side-effects of a mitigation strategy are not clearly understood by a security expert, the consequences may be costly. Thus, vulnerability mitigation requires a deep understanding of the subtle interdependencies that exist between domains that are different in nature. This is especially difficult for new technology use models, such as Cloud-based computing and IoT, in which cyber and physical components are combined and interdependent. By their own design, ontologies and the associated inference mechanisms permit us to reason about connections between diverse domains and contexts that are pertinent for the general threat picture, and to highlight the effects and ramifications of the mitigation strategies considered. In this paper, we position ontologies as crucial tools for understanding the threat space for new technology space, for increasing security experts' situational awareness, and, ultimately, as decision-support tools for rapid development of mitigation strategies. We follow with the discussion of the new information and insights gleaned from the ontology-based study of the root of trust in cyber-physical systems.

1 Introduction

Modern processes and technologies are cross-domain, merging together approaches created for different contexts. Complexity is intrinsic. Even activities resulting in identical or similar outcomes – e.g., sending electronic mail, processing identical datasets or payments, using e-commerce applications, or assessing the data quality collected from sensors – could be executed in very different environments, resulting in different risks. Thus, it is sometimes necessary to assume different risk postures in response to similar events or in the course of the same process.

Moreover, computing or physical environments are not the only contexts that influence the nature of vulnerabilities. Economic conditions or regulatory requirements can alter the impact of the cybersecurity risks and therefore lead to changes in mitigation strategies.

This paper uses cyber-physical systems (CPS) as an example environment. Complexity and composition considerations are especially meaningful when analyzing CPS that have computing capabilities, communication (cyber) capabilities, and physical interfaces [1]. In most use cases, CPS and other systems don't operate in isolation, but rather work in the end to end continuum, extending from edge devices to the Cloud, where data generated by sensors and enriched by processing can be stored. CPS almost always display significant environmental complexity as do multi-device environments in general, complicated by the physical interfaces and use cases that CPS generally enable. Moreover, diversity among CPS is extensive, with seemingly little in common in different CPS contexts. If we compare connected kitchen appliances with transportation systems, or energy systems, they appear to have very little in common, but they draw from similar foundational technologies and deployment processes.

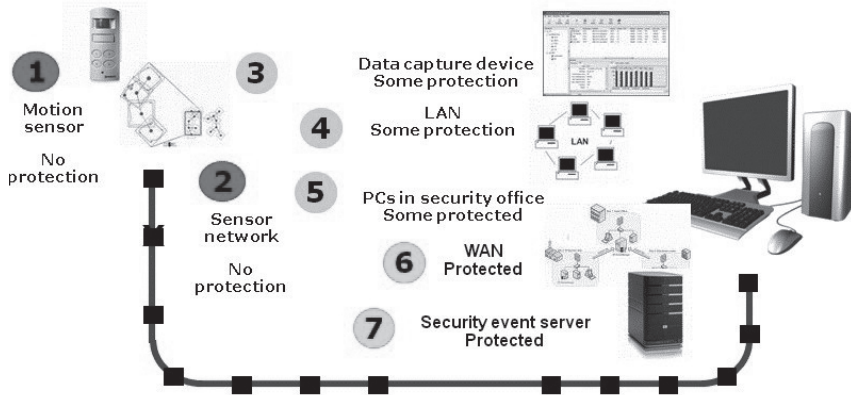


Fig. 1: Security protection and “anonymity readiness” in today’s computing environments.

Because of the complexity of processes associated with the use of CPS, one process typically contains multiple operational conditions and levels of security capabilities as illustrated in Fig. 1.

Although the complex processes have a unifying operational goal, the security capabilities are different at different stages of the process. The diversity in security extends to privacy protection.

The diversity and complexity makes it impossible to assess composite risks with traditional techniques [GGIK2015] and to develop mitigations that are broadly applicable rather than context dependent. Ontology-based reasoning can permit us to obtain a multi-dimensional view of the subject, incorporate consistent constraints, understand dependencies, and make informed conclusions about remediation. It could also help the developers to design a nuanced risk posture in new technologies that is better suited to the majority of today’s dynamic use cases. Finally, we believe ontologies could be useful in assessing new and emerging technology spaces, for both research and technology deployment, in multi-disciplinary subjects like cybersecurity.

2 Ontology and Complex Multidisciplinary Subjects

2.1 About cybersecurity

Cybersecurity has begun to crystallize into a firmer subject a relatively short time ago. Although definitions of cybersecurity vary, they are not highly divergent and frequently comprise a narrow definition and a broader one. The example of a narrow definition is provided by the National Initiative on Cybersecurity in the US:

The activity or process, ability or capability, or state whereby information and communications systems and the information contained therein are protected from and/or defended against damage, unauthorized use or modification, or exploitation.

In other cases, a broad definition, including related and non-technical subjects, from economics and psychology to political science and diplomacy is used, for example:

Strategy, policy, and standards regarding the security of and operations in cyberspace, and encompassing the full range of threat reduction, vulnerability reduction, prevention, international engagement, incident response, resiliency, and recovery policies and activities, including computer network operations, information assurance, law enforcement, diplomacy, military, and intelligence missions as they relate to the security and stability of the global information and communications infrastructure.

Further complicating the issue, cybersecurity characteristics for IT systems, the best studied cybersecurity area, are different from the emerging characteristics of cybersecurity when applied to different environments, such as cyber-physical systems (CPS). While the researchers have defined cross-cutting considerations that apply to most cybersecurity environments, the science of cybersecurity is not sufficiently advanced to create a unifying theory of cybersecurity foundations. As a result, every new context (such as CPS) tends to develop its own cybersecurity approach that shares similar technologies and governance models with adjacent contexts, but creates its own body of knowledge. Cybersecurity approaches for energy sector (e.g., smart meters) differ superficially from the approaches adopted in transportation (e.g., smart cars), leading to the fragmentation of cybersecurity and slower adoption of productive techniques. In new areas, the technology community tends to focus on niche context driven issues because they are easier to analyze and to avoid studying broadly applicable phenomena. Consequently, literature describing cybersecurity R&D in energy space draws very little content and engages in little collaboration with the researchers focusing on transportation, although technologies used in both contexts are very similar.

An ontology-based approach could permit researchers and practitioners to link together disparate content that draws from similar premises [IBN+15], allowing technologists to reuse, share, and propagate knowledge, in order to create a field of cybersecurity that is broader in scope and more theoretically sound.

2.2 About ontology

One of the main goals of the field of Knowledge Representation (KR) is the study of methodologies and tools that enable capturing knowledge accurately, compactly and so that information

can be easily added and updated. Acknowledging the importance of the design principle of separation of concerns, KR researchers typically separate knowledge specification from the associated computations. This yields a *declarative specification* (as opposed to the traditional *imperative* one), in which knowledge is specified by statements that say what is true (or false) in the domain of interest, without stating how, algorithmically, statements should be combined and their truth propagated. Rather, the semantics of the representation language defines the meaning of those statements in precise and unambiguous (usually, logical) terms. For automated computation, general-purpose algorithms, often called *inference engines*, are separately defined, which embody the semantics of the language. Thanks to this separation, the meaning of a *knowledge base* can be determined independently of the particular algorithms used, and alternative algorithms can be adopted to fit specific practical needs (e.g., performance on given kinds of knowledge).

An ontology is a hierarchical specification of a set of objects from a domain of interest, of their properties and of their relationships. As such, ontologies enable a principled organization of knowledge. For example, a simple ontology may specify that laptops and desktops are kinds of computers, that computers and smartphones are kinds of computing devices, that all computing devices are equipped with a CPU, and that computers and smartphones are disjoint classes of objects (i.e., something cannot be a computer and a smartphone at the same time). Additionally, the ontology may specify that “John’s workstation” is a laptop. Specifically designed ontological languages enable the encoding of such knowledge in an accurate way.

The true power of ontologies, however, comes from the fact that ontological languages are associated, through their semantics, with inference mechanisms that make it possible to perform automated, provably correct reasoning about the elements of an ontology. Inference mechanisms are related, for example, to expanding the class-subclass relationships into ancestor/descendant and – importantly – to determining how properties and relationships are propagated through the hierarchy specified by the ontology, i.e., how classes inherit their ancestors’ characteristics. In the computer ontology described earlier, inference mechanisms can conclude, for example, that laptops are computing devices and that, as such, they inherit the properties of the latter. Hence, it is possible to infer that all laptops are equipped with a CPU and that “John’s workstation” is equipped with a CPU. Finally, because computers and smartphones are disjoint classes, it is possible to conclude that “John’s workstation” cannot be classified as a smartphone.

By applying inference mechanisms, one can often derive information that was not immediately evident from the original specification of the ontology, and the reasons for such derivation can be clearly pinpointed and explained automatically.

Notable similarities exist between ontologies and (relational) databases, which in fact can be viewed as their precursors. Like ontologies, databases are declarative specifications of objects and of their properties and relations. From a conceptual perspective, however, ontologies are characterized by a more uniform and thorough encoding of knowledge. For example, information about computing devices can indeed be encoded using a relational database, but the meaning (i.e., the semantics) of the relations themselves remains implicit and external to the database. Thus, while the database may well contain a relation (represented by a table) called “kind-of” that holds between laptop and computer, the meaning of such relation – e.g., its transitivity and the inheritance of properties from classes to their sub-classes – is not part of the specification and must be provided separately to draw inferences.

2.3 Reasoning about multidisciplinary connections using ontologies

The general-purpose, hierarchical nature of ontologies, their broad applicability, and the fact that all relevant information is encoded in an explicit, machine-accessible way, make ontologies prime candidates for formalizing multidisciplinary knowledge and for reasoning about the underlying connections.

An interesting example of a multidisciplinary ontology is that of [PFCS14], in which a multidisciplinary ontology of epidemiology is developed in order to enable a uniform annotation of epidemiology resources and the integration and sharing of data about global epidemiological events.

A further example is that of [BaLR14], where the authors discuss how an ontology could be used for the development of a discovery network linking databases of materials scientific data. Such a “Layered Material Ontology” would enable connecting multidisciplinary knowledge ranging from matter and materials to performance and design (see Fig. 2), and asking queries spanning across domain boundaries, such as asking for metal alloys that are suitable for a given kind of design.

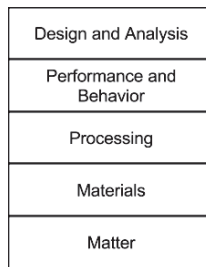


Fig. 2: Structure of the Layered Material Ontology proposed (courtesy of [BaLR14]).

From an organizational perspective, when tackling multidisciplinary knowledge, it is useful to divide the formalization in upper ontology and (multiple) domain ontologies. An upper ontology is an encoding of the concepts that are common across all disciplines of interest. In the context of securing cyber-physical systems, for instance, an upper ontology might define the high-level concept of system component, with its refinements of computational device and physical device, and the concept of vulnerability. Additionally, a relation “vulnerable-to” might be used to associate a system component with its known vulnerabilities.

A domain ontology formalizes a specific knowledge domain. The concepts captured by a domain ontology are typically specified as specializations of concepts from the upper ontology. In reference to the previous example, a domain ontology of smart grids might describe SCADA systems as kinds of computational devices, power generators as types of physical device, and list a number of vulnerabilities specific to the smart grid. Relation “vulnerable-to” could then be used to indicate the specific vulnerabilities of smart grid components. Similarly, a domain ontology of automotive systems might describe the ECU as a computational device, a brake actuator as a physical device, and use relation “vulnerable-to” to specify the vulnerabilities of the various components of an automotive system. Inference can then be applied to propagate relevant properties and relations throughout the ontology. For example, suppose a new vulnerability is discovered, which

affects certain system components. Obviously, one can check which components are directly vulnerable. However, one may also want to define the notion of a component being “affected by” the vulnerability either because it is directly vulnerable to it or because it is connected to some other component that is affected by it. Inference can now be used to identify, across the ontology, any component that is affected, even remotely, by a vulnerability (more details on this topic will be provided in the next section).

This representation and reasoning framework becomes especially useful in situations in which knowledge from multiple fields must be taken into account at the same time. Consider the task of assessing the vulnerabilities of an electric car. The example ontology discussed in the previous paragraph would allow one to study vulnerabilities that may come from coordinated exploits affecting both the power system and the braking system (e.g., one could automate the search for scenarios in which a central control component becomes overloaded when elements in the power *and* braking subsystems are caused to misbehave). Such a model can be incrementally extended by adding domain ontologies for other car subsystems. By replacing the braking system ontology with an ontology modeling a navigation system or a weapons system, one could study the vulnerabilities of combat ships. What is essential to note is that, in all of these cases, multidisciplinary knowledge can be incrementally and seamlessly integrated and sophisticated questions about the systems being modeled can be answered by means of general-purpose inference mechanisms, without the need to develop dedicated algorithms.

3 Case Study: Root of Trust in CPS

3.1 General information about the project

The Cyber Security Research Alliance, Inc. (CSRA) is an industry-led, non-profit consortium focused on research and development strategy to address evolving cyber security environment through partnerships among government, industry, and academia. This effort was established in response to the growing need for increased public-private collaboration to address R&D issues in cyber security.

CSRA has identified several priority areas crucial for improving security in cyber-physical systems through the input at the CSRA/NIST Workshop on Cybersecurity for Cyber Physical Systems, held April 4-5, 2013. Almost all of the study sessions acknowledged the need for the common vocabulary and reasoning mechanism to unify currently available research and technology to reduce fragmentation of CPS space. The lack of common terminology and combined assessment of work in adjacent fields was considered one of the main inhibitors of research due to the diversity of CPS contexts and the multidisciplinary nature of the field. As a result, best practices and research advances are not always shared and applied across relevant CPS contexts.

Following the workshop, CSRA set up a pilot project to build a subset of an ontology focusing on cybersecurity for CPS. The project covered the subject of the root of trust in CPS. Teams from two universities – George Mason University and Drexel University – participated in the pilot and built the foundations of the ontology.

The participants in the pilot project surveyed the field, prioritized technologies, identified gaps, and defined ontology approaches that could be adjusted for CPS contexts. Seed ontologies were

built by both groups of the participants. They included terminology, information on research already done, R&D groups active in this area, and other relevant information. The reasoning process identified and prioritized gaps that need to be addressed. The project addressed the discovery of cyber security technologies protecting CPS at different stages of development. The results of the project have been used as a tool for subsequent phases of research in COS security addressing research gaps, evaluating research results, directions of technology adoption and commercialization, e.g., as a reference point in the work of the NIST Public Working Group (PWG) on cyber-physical systems. The outcome of the project helped multidisciplinary teams investigate solutions in perspective of real-world trade-offs for protection, detection and response to cyber-attacks on CPS.

We provide information on one of the project deliverables below.

3.2 Building the ontology

The RoT ontology is divided into an upper component, which provides concepts relevant to all cyber-physical systems, and domain ontologies for the specific domains, including smart grids, transportation, and healthcare. Key elements of the upper component are the notions of *cyber-physical system concepts*, *cyber attacks*, and *countermeasures*. The latter two classes are divided in further domain-independent concepts, such as *malicious* and *non-malicious threats* and *cyber defense methods* (e.g., *preparation* and *detection*). Although there is an obvious relationship among the three top components of the upper ontology, to ensure breadth of the ontology we have included in it elements as exhaustively as possible, independently of whether they are currently related to other elements from the ontology. For example, instances of cyber attacks have been included independently of whether it is currently known how to use them against cyber-physical system. Domain ontologies provide further specializations of the three top components. Next, we focus on the smart-grid domain ontology, SG.

The development of the SG ontology was guided by the principles outlined in [LNB+15]. Information was obtained from subject matter experts and from various published sources, including [WaLu13], [NIST10], and [CMGS12]. Fig. 3 gives an overview of the upper component and of the SG domain ontology.

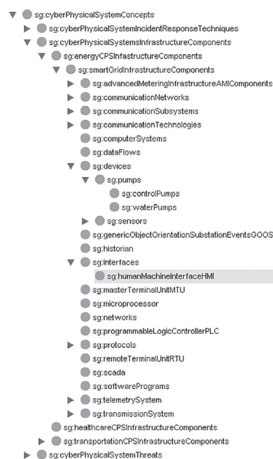


Fig. 3: RoT ontology – upper component and SG domain ontology

At the root of the SG ontology is the concept of *energyCPSInfrastructureComponents*, which acts as a superclass of any concept related to the energy CPS infrastructure and enables expanding the ontology to other energy cyber-physical systems beyond smart grids. Directly under it is class *smartGridInfrastructureComponents*, which constitutes the root concept for smart-grid components. The organization of its subclasses follows an organizational paradigm that is intended to be applicable, with relatively small changes, to multiple knowledge domains. According to this paradigm, concepts are classified in one of:

- *devices*
- *interfaces*
- *protocols*

For example, in the case of our smart grid domain, the class of *devices* comprises *sensors* and *pumps*. Other notable subclasses include:

- *scada*
- *historian*
- *masterTerminalUnitMTU*
- *remoteTerminalUnitRTU*

All of these classes represent key devices of the smart grid infrastructure; the SCADA system, for example, acts as a central governor of the infrastructure, communicating with, and controlling, all remote equipment. There are also classes for key subsystems, such as *telemetrySystem* and *transmissionSystem*.

Property	Value
rdf:type	owl:Class
rdfs:subClassOf	sg:interfaces
sg:trusts	sg:masterTerminalUnitMTU
sg:vulnerable_to	sg:bufferFlooding

Fig. 4: Class humanMachineInterfaceHMI and relations trusts and vulnerable_to

The most fundamental relation defined by the ontology is the trust relationship, informally denoting the fact that one component trusts another. Intuitively, if a trusted component is affected by a cyber threat, the trusting component will also likely be affected, either directly (e.g., by being compromised) or indirectly (e.g., because it takes as credible false information that is fed to it from the affected component). The fact that a component is vulnerable to a certain threat is encoded by relation *vulnerable_to*. Fig. 4 shows a sample class and the corresponding definitions of the relations. More specifically, we see that human-machine interface (HMI) “trusts” the master terminal unit and that the HMI is vulnerable to buffer flooding.

3.3 Insights obtained from the ontology

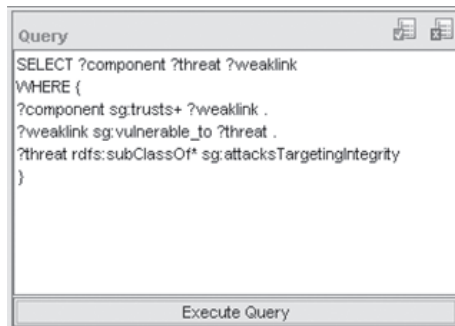
The RoT ontology makes it possible to answer a number of important and nuanced questions related to the assessment of the weaknesses of the infrastructure, including:

- What elements does a given component (e.g., SCADA) trust?
- In turn, what do these elements trust?

- What vulnerabilities does this component have?
- What is impacted by a given vulnerability?

Although conceptually simple, these questions involve a rather substantial amount of reasoning. Consider for instance the last question, useful in a scenario in which a vulnerability is discovered and one wants to determine all components that are put in danger by this vulnerability. Generally speaking, the components that are directly affected by the vulnerability are to be identified (using relation *vulnerable_to*), and then the information must be propagated recursively (through relation *trusts*) to all components trusting the vulnerable ones either directly or indirectly. If the infrastructure includes only a small number of components, then the answer may be straightforward. However, in larger infrastructures answering the question may be more challenging due to more complex trust chains.

With traditional approaches, answering these questions would likely involve implementing a different algorithm for each of them, algorithms (and corresponding data structures) that are made non-trivial by the variability of the concepts that need to be represented. The adoption of an ontology-based formalization makes it possible to accomplish all of this by stating the questions in a declarative fashion (i.e., by specifying what one is looking for, rather than how to find it) and without the need for implementing ad-hoc algorithms. This is achieved thanks to the general-purpose inference mechanisms associated with the ontology and to powerful query languages. For instance, the components that may be affected by an attack targeting integrity can be found by means of the query shown in Fig. 5.



```
Query
SELECT ?component ?threat ?weaklink
WHERE {
  ?component sg:trusts+ ?weaklink .
  ?weaklink sg:vulnerable_to ?threat .
  ?threat rdfs:subClassOf* sg:attacksTargetingIntegrity
}
Execute Query
```

Fig. 5: Sample query

Intuitively, the query asks the inference mechanism to find all triples of the form such that is a “trust element” of , and it is vulnerable to , where is, in this example, a type of attack targeting integrity. Similar queries allow one to identify all trust elements of a given component and to determine the starting points of the corresponding trust chains, which can be viewed as the *roots of trust*.

Fig. 6 shows the output of a query requesting the trust elements of SCADA.

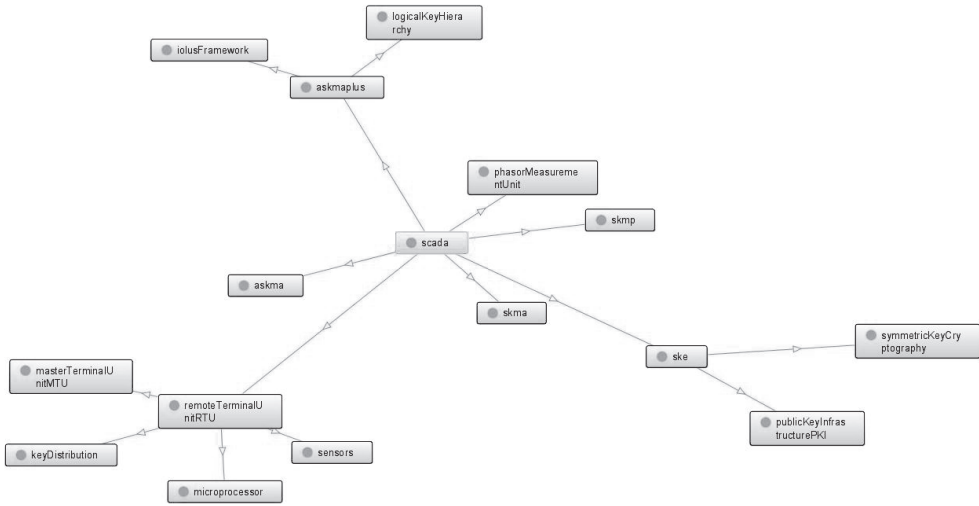


Fig. 6: Trust elements of SCADA

4 Conclusions and Future Work

The review of research literature focusing on cybersecurity in different CPS contexts shows commonalities in approaches among different types of systems, although collaborations among scientists focusing on different contexts remains minimal. Similarly, interactions between different domains in cybersecurity, e.g., safety, reliability, and security proper, are limited, although they are reflected in recent literature, e.g., [SMSG15]. Various approaches have been tried to facilitate greater flow of ideas among different contexts and enhance multidisciplinary collaboration, but the nature of synergies remains difficult to assess, and the results difficult to evaluate.

We believe ontological reasoning could be instrumental in fostering a consistent emerging technology space, helping realize broadly applicable ideas in a field of research, and maximize the ability to bring these ideas to practice.

The pilot project for CPS root of trust helped the research teams to identify approaches to creating knowledge representations for a specific, but complex field. The tools created as a result assisted the research community and practitioners to form a multi-dimensional view of emerging subjects, identifying gaps, priorities, and affinities with adjacent fields¹.

The project paved the way for continued research in ontologies for emerging fields. Further work will include broader analysis with a larger number of contexts as well as the creation of focused analysis tools specialized for R&D, research funding, or deployment in new areas of technology.

¹ See http://www.cybersecurityresearch.org/news_and_events/press_releases/pr_20150106.html for more information.

Bibliography

- [FKWL11] Ashfaq H. Farooqi, Farrukh A. Khan, Jin Wang, and Sungyoung Lee. 2011. Security requirements for a cyber physical community system: a case study. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*(ISABEL ,11). ACM, New York, NY, USA
- [GGIK15] Dieter Gollmann, Pavel Gurikov, Alexander Isakov, Marina Krotofil, Jason Larsen, and Alexander Winnicki. 2015. Cyber-Physical Systems Security: Experimental Analysis of a Vinyl Acetate Monomer Plant. In *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*(CPSS ,15). ACM, New York, NY, USA, 1-12.
- [BaLR14] Balduccini, M., LeBlanc, E., & Regli, W. C.: Towards a Content-Based Material Science Discovery Network. In: *2014 AAAI Workshop for Discovery Informatics*, 2014.
- [CMGS12] Cichonski, P., Millar, T., Grance, T., and Scarfone, K.: Computer Security Incident Handling Guide. National Institute of Standards and Technology Special Publication No. 800-61, 2012.
- [IBN+15] Iannacone, Michael, Bohn, Shawn, Nakamura, Grant, Gerth, John, Huffer, Kelly, Bridges, Robert, Ferragut, Erik, and Goodall, John. 2015. Developing an Ontology for Cyber Security Knowledge Graphs. In *Proceedings of the 10th Annual Cyber and Information Security Research Conference*(CISR ,15). ACM, New York, NY, USA
- [LNB+15] LeBlanc, E., Nguyen, D., Balduccini, M., Regli, W. C., Kopena, J., and Wambold, T.: Military Ontologies for Information Dissemination at the Tactical Edge. In: *1st Workshop on Formal Ontologies for Artificial Intelligence (FOfAI15)*, 2015.
- [NIST10] National Institute of Standards and Technology. *Guidelines for Smart Grid Cyber Security. NIST IR-7628, 2010.*
- [PFCS14] Pesquita, C., Ferreira, J. D., Couto, F. M., and Silva, M. J.: The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources. In: *J. Biomed Semantics*, 5(4), 2014.
- [WaLu13] Wang, W., and Lu, Z.: Cyber Security in the Smart Grid: Survey and Challenges. In: *Computer Networks*, 57(5), 2013, 1344-1371.
- [SMMSG15] Christoph Schmittner, Zhendong Ma, Erwin Schoitsch, and Thomas Gruber. 2015. A Case Study of FMVEA and CHASSIS as Safety and Security Co-Analysis Method for Automotive Cyber-physical Systems. In *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*(CPSS ,15). ACM, New York, NY, USA, 69-80.