An Ontology for Insider Threat Indicators

Development and Applications

Daniel L. Costa, Matthew L. Collins, Samuel J. Perl, Michael J. Albrethsen, George J. Silowash, Derrick L. Spooner
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA, USA
insider-threat-feedback@cert.org

***Abstract*—We describe our ongoing development of an insider threat indicator ontology. Our ontology is intended to serve as a standardized expression method for potential indicators of malicious insider activity, as well as a formalization of much of our team's research on insider threat detection, prevention, and mitigation. This ontology bridges the gap between natural language descriptions of malicious insiders, malicious insider activity, and machine-generated data that analysts and investigators use to detect behavioral and technical observables of insider activity. The ontology provides a mechanism for sharing and testing indicators of insider threat across multiple participants without compromising organization-sensitive data, thereby enhancing the data fusion and information sharing capabilities of the insider threat detection domain.**

Keywords—ontology; insider threat; data fusion; information sharing

I. Background

The study of insider threat presents some of the most complex challenges in information security. Even defining the insider threat has proven difficult, with interpretations and scope varying depending on the problem space. The CERT® Division of Carnegie Mellon University's Software Engineering Institute defines a malicious insider as a current or former employee, contractor, or other business partner who has or had authorized access to an organization's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems [1]. Organizations have begun to acknowledge the importance of detecting and preventing insider threats, but there is a surprising lack of standards within the insider threat domain to assist in the development, description, testing, and sharing of these techniques. For many organizations, establishing an insider threat program and beginning to look for potentially malicious insider activity is a new business activity. In particular, Executive Order 13587 and the National Insider Threat Policy describe minimum standards for establishing an insider threat program and monitoring employee use of classified networks for malicious activity [2-4].

II. Purpose

A. Goals

The primary goal of this effort is to support the creation, sharing, and analysis of indicators of insider threat. Because insider data is sensitive, insider threat teams often work only with data from inside their own organizations. These records frequently include documented employee behaviors, intellectual property, employee activity on networks, and information on organizational proprietary networks and information technology (IT) architecture. Organizations and teams are hesitant to release this information due to the risk of breaching employee privacy, releasing sensitive organizational information, or unnecessarily losing a competitive advantage. A shared ontology will allow teams to share indicators of insider threat without disclosing their own sensitive data. Our desired outcome is to facilitate information sharing on effective indicators of malicious insider activity across organizations, with an emphasis on extensibility, semi-automation, and the ability for community members to benefit from investigations and analysis performed by others.

B. The Case for an Ontology

All entity and relationship data models, including semantic data models, have their limitations [5]. Models are extremely formal by design and can encounter problems when representing the variety of actions involved in a real-world insider threat case. In addition, the data on cases of insider threat is often gathered from legal judgments and outcomes whose documentation is highly variable. As a result, insider threat domain experts tend to rely on natural language to document their cases and findings. Though natural language is more expressive than a model, we believe the insider threat domain will benefit from the development of an ontology. Our interest in building an ontology, developed from our observations of the field today, is driven by the following factors:

- We expect rapid growth in the data being collected and shared by organizations, specifically about insider threats. Some organizations have already stated that overcoming this challenge is one of their top priorities [6].
- The insider threat research community lacks a defined, formal model that is machine readable, human understandable, and transferrable with limited sharing

barriers. We felt that starting a model of this kind, based on the real-world case data we have already collected, could accelerate this process within the community, as has been done in other fields [7, 8].

- We are willing to accept some loss of descriptive power for individual cases, provided we can analyze large populations of cases using computation. We expect insider threat teams (both in research and in operations) to be asked to detect insider threat activity by analyzing a growing quantity of data from new sources in an increasingly limited amount of time.

III. Approach

A. Domain Identification

At first glance, defining the domain of our ontology appeared to be a trivial matter: representation of potential indicators of malicious insider activity. In practice, indicators of malicious insider activity involve complex interconnections of parts of several other domains:

- Human behavior: understanding insider threats involves understanding the people behind the malicious activity—the reasons why they attacked, their psychological characteristics, their emotions, and their intent.
- Social interactions and interpersonal relationships: modeling the relationships between insiders and their employers, colleagues, friends, and family is a crucial part of identifying stressors that are often associated with malicious insider activity.
- Organizations and organizational environments: the culture and policies of organizations factor heavily into the interpretation of malicious behavior within an organization.
- Information technology security: information and information systems can be both the targets of and tools used to perpetrate malicious insider activity. IT security also contains other concepts of interest in describing the insider threat domain, namely, confidentiality, integrity, and availability.

B. Domain Scoping

With a representative list of sub-domains for insider threat enumerated, our next challenge was determining the scope at which our ontology must provide support for each subdomain. We chose to develop the following competency questions for our ontology to assist us in our scoping efforts [9, 10].

- What concepts and relationships comprise the technical and behavioral observables of potential indicators of malicious insider activity?
- What potential indicators of malicious insider threat activity are insider threat teams using for detection?
- To facilitate information sharing, at what level of detail should organizations describe their indicators of malicious insider activity without revealing organization-sensitive information?

C. Construction Method

Since 2001, the CERT® Insider Threat Center has collected over 800 cases in which insiders used IT to disrupt an organization's critical IT services, commit fraud against an organization, steal intellectual property, or conduct national security espionage, sabotaging systems and data, as well as other cases of insiders using IT in a way that should have been a concern to an organization. This data provides the foundation for all of our insider threat research, our insider threat lab, insider threat assessments, workshops, exercises, and the models developed to describe how the crimes evolve over time. Our case collection involves gathering and analyzing data from public (e.g., media reports, court documents, and other publications) and nonpublic (e.g., law enforcement investigations, internal investigations from other organizations, interviews with victim organizations, and interviews with convicted insiders) sources. This data collection, summarized in Figure 1, primarily focuses on gathering information about three entities: the organizations involved, the perpetrator of the malicious activity, and the details of the incident. Each case in our insider incident repository contains a natural language description of the technical and behavioral observables of the incident. We used these descriptions as the primary data source for our ontology.

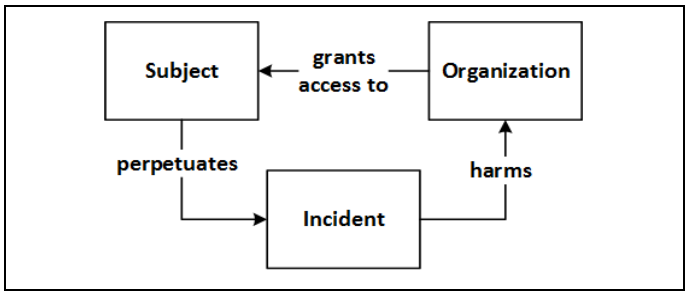

Fig. 1. CERT model for insider incidents

1) Data-Driven Ontology Bootstrapping

To ensure full coverage of the information contained in our insider incident repository, we adopted an approach that utilizes concept maps as a first step in the development of an ontology [11]. Manually developing concept maps for over 800 individual insider threat cases required an infeasible level of effort, so we developed a semi-automated concept map extraction method adapted from several existing approaches [12, 13]. This method used part-of-speech and part-of-sentence tagging to extract [concept, concept, relationship] triples from the natural language description of each insider incident. We utilized additional text and natural language processing techniques to eliminate stop-words, group similar triples, and sort the triple collection by frequency of occurrence. We then used this collection of triples as the basis for our class hierarchy, using our competency questions to set scope and optimize the arrangement of specific classes.

2) Additional Data Sources

We supplemented the candidate classes and object properties derived from our insider incident repository with concepts and relations from the cyber threat and digital forensics domains. We reviewed the Structured Threat

Information Exchange (STIX) and Cyber Observable Expression (CybOX) languages [14, 15], as well the SANS Institute's digital forensics artifact catalog [16], to fill gaps in our concepts for cyber threats, cyber observables, and their associated forensic artifacts.

IV. Implementation

A. Design Decisions

We adapted components from several existing ontologies for our work. To assist in the modeling of actors and their actions, we adapted several top-level ontology components from material available on schema.org [17]. We leveraged existing ontologies for filling gaps in our coverage of cyber assets, including concepts from the network services, IT systems, IT security, and mobile device domains [18-21]. To validate our design, we used the catalog of common ontology development pitfalls from work titled "Validating ontologies with oops!" [22]. We provided support for modeling the temporality of actions and events relative to one another through use of the sequence design pattern [23]. We have chosen to implement our ontology using the Web Ontology Language (OWL), due to its maturity, wide use, and extensibility [24].

B. Overview of Top-Level Classes

The top-level of our ontology, summarized in Figure 2, is composed of five classes: Actor, Action, Asset, Event, and Information. The Actor class contains subclasses for representing people, organizations, and organizational components such as departments. The Action class contains the subclasses that define the things that actors can perform. The Asset class provides subclasses that define the objects of actions. The Information class provides subclasses that provide support for modeling the information contained within some assets (examples include personally identifiable information, trade secrets, and classified information). The Event class provides support for multiple types of events of interest. Events are generally associated with one or more Actions. The creation of an individual event typically requires making some inference, as opposed to an individual Action, which can be created through direct observation. For example, moving a file is modeled in our ontology as an Action. A data exfiltration event, when associated with a file move action via the *hasAction* object property, expresses the fact that the associated action was unauthorized. Additionally, an object property hierarchy is provided to express various types of relationship roles, job roles, and event roles.

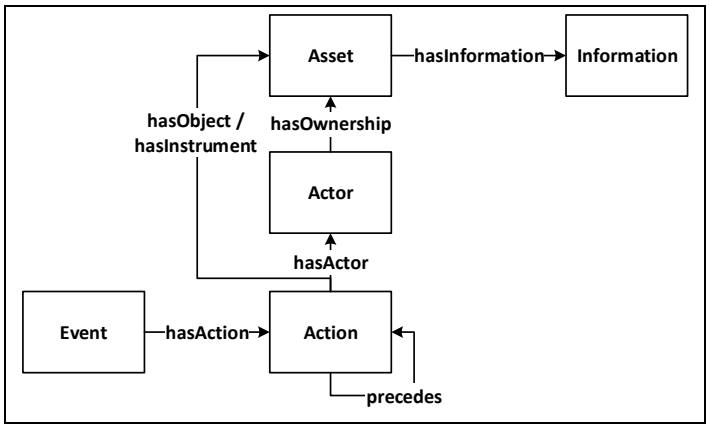

Fig. 2. Top-level ontology classes and object properties

C. Example Uses

To demonstrate use of the ontology to describe indicators of malicious insider activity, we present two examples of translating natural language descriptions of indicators of malicious insider activity from our insider threat incident repository into ontology individuals. The translation process is relatively straightforward; the concepts from each description are manually identified, individuals are created for each concept as instances of the appropriate ontology class, and individual object properties are added to relate the class instances to one another. Figure 3 and Figure 4, respectively, depict the ontology translation for the following insider threat indicator descriptions:

- The insider transferred proprietary engineering plans from the victim organization's computer systems to his new employer.
- The insider accessed a web server with an administrator account and deleted approximately 1,000 files.

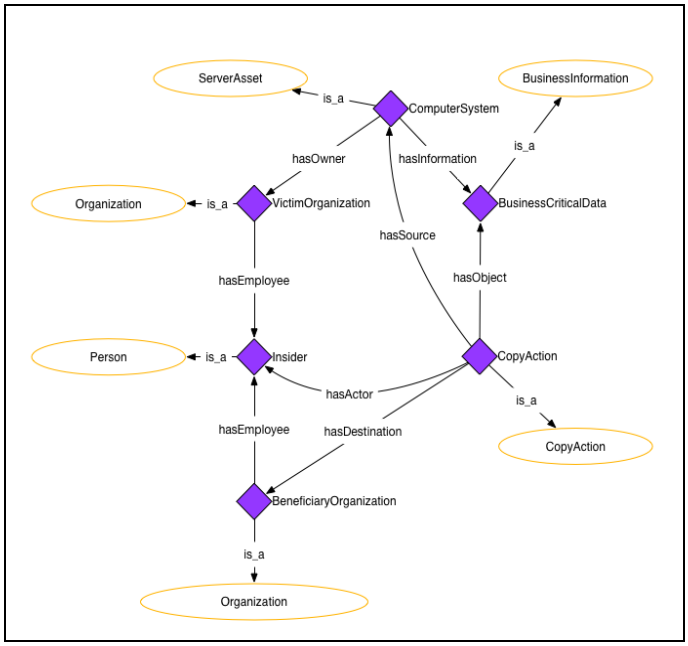

Fig. 3. Data exfiltration example from insider incident repository translated into ontology individuals

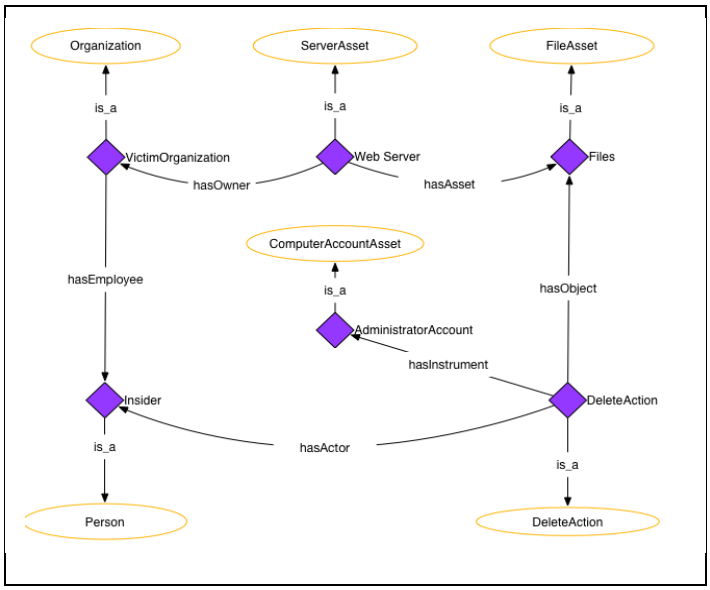

Fig. 4. Information technology sabotage example from insider incident translated into ontology individuals

V. APPLICATIONS

A. *Insider Threat Indicator Information Sharing*

Our ontology provides two powerful concepts in the description of potential indicators of malicious insider activity: abstraction and extensibility. By abstraction, we mean that indicators can now be described at a level of detail that omits organization-sensitive information while still maintaining enough descriptive information to express the idea that given observable actions or conditions are potential indicators of malicious insider activity. By extensibility, we mean that we have provided the conceptual components that organizations can use to describe their existing indicators and develop new indicators. Potential indicators of malicious insider activity often include qualifiers such as "excessive," "anomalous," "unauthorized," and "suspicious" to distinguish conditions that are potentially indicative of malicious insider activity from "normal" behavior and activity. Definitions and interpretations of these types of conceptual qualifiers vary greatly from organization to organization, and often vary within organizations based on variables such as job type, location, and time. To accommodate these variations, we introduce the idea of "policy packs" in our ontology: modular collections of ontology axioms that represent organization-agnostic concepts, definitions, and interpretations of indicator patterns. Our ontology specifically provides support for this via the Event class hierarchy. Organizations using our ontology can develop their own defined classes, or modify existing ones, to specify the necessary and sufficient restrictions for class membership.

B. *Automated Indicator Instance Extraction Framework*

Insider threats can be detected by observing instances of indicators of malicious insider activity within an organization. Operationally, this involves the collection and analysis of large amounts of data on every employee in an organization. Without some level of automation, this detection practice becomes infeasible to perform effectively and efficiently. Using our ontology, we have designed a semi-automated approach for the detection of potential indicators of malicious insider activity that fuses data from multiple types of sources. The ontology provides an analysis hub that combines information from an organization's enterprise network activity and human resources data to provide a data-rich environment for the development and detection of robust, effective indicators of malicious insider activity.

1) *Operational Data to Ontology Individuals*

We use the term "operational data" to encapsulate the data and data sources that capture the user-based activity that occurs on an organization's information systems and networks. The technical observables associated with some potential indicators of malicious insider activity are found in operational data and during the analysis of trends in operational data. Some examples of operational data include:

- Host-based user activity logs
- Critical application audit logs
- Network activity logs
- Communication server logs
- System event logs

Since operational data is usually found in structured or semi-structured log files, we attempted to prove the concept of automatically translating the information contained in operational data sources into ontology individuals. Instead of direct translation into ontology individuals from operational data sources, we chose to translate the operational data into CybOX cyber observable files, and automatically create ontology individuals based on the contents of the CybOX files. This approach allowed us to focus on identifying the fields from CybOX that were applicable to our ontology classes, and provide a translation mechanism for only those applicable fields. Without the CybOX translation layer, we would have had to develop ontology translation mechanisms for each type of operational data source we wish to support, which would require an infeasible level of effort, support, and maintenance. Additionally, CybOX provides an API for their XML file format, which facilitates the automated translation of any input data source into the CybOX format. (CybOX currently supports over 60 input data sources.)

In our proof of concept, we were successful in automatically translating Windows system event logs into the CybOX format, and, using simple scripts, automatically generating the OWL XML code to create individuals for a small subset of our ontology classes. In a robust implementation, the automated ontology individual creation would provide configurable settings that would allow organizations to control the creation of ontology individuals for classes whose specific definitions may vary from organization to organization. For example, if the ontology contained a class representing after-hours logins, the automated individual creation mechanism should provide a way to specify a time range that is considered after-hours.

2) Human Resources Data to Ontology Individuals

We use the term "human resources data" to encapsulate data and data sources that provide contextual and behavioral information about employees. These records are typically stored in an unstructured format, and are locked within Human Resources departments to protect the privacy rights of employees. Examples of human resources data include:

- Organization charts
- Employee performance reviews
- Employee personnel files, including job title, supervisor, role, and responsibilities
- Employee behavior records, such as formal reprimands and policy violations
- Information from anonymous insider reporting channels
- Results of background checks

Human resources data provides a rich source of contextual, behavioral, and psychosocial information regarding employees. Human resources data is typically more fragmented and less structured than operational data, so the automated translation of this data into ontology individuals may be a challenge for some organizations. Enterprise solutions for human resource information management exist, and where they are used, a structured representation of human resources data could be used to develop an automated ontology translation process. In our proof of concept for the automated indicator instance extraction framework, we did not attempt to automatically create ontology individuals from human resources data, but in future work, we will apply a similar approach to we used for operational data.

3) Semantic Reasoner

If operational data and human resources data are both described using the ontology, and if indicator policy packs are in place, an organization can use a semantic reasoner to make inferences and automatically classify ontology individuals as instances of specific defined classes. Ontology individuals that meet the formal definitions of potential indicators of malicious insider activity can then be said to have "satisfied" some indicator. A collection of ontology individuals that satisfy threat indicators becomes a useful data set for insider threat analysts. With a robust set of indicators implemented as defined classes, analysts have the ability to see descriptions of potential indicators of malicious insider activity across previously disparate data sets and at larger scale. Satisfied indicators can be reviewed by analysts to identify false positives, refine indicators, develop new indicators to add back into the ontology via policy packs, or create threat reports that summarize the potential malicious insider activity found in the data.

4) Putting it All Together

The full framework—beginning with the development and maintenance of the ontology through the release of organizational threat reports based on the detected instances of potential indicators of malicious insider activity—is presented in Figure 5. This framework is meant to support detection of potential indicators of malicious insider activity that is then triaged. An effective implementation of the framework depends on the indicators it contains, and not all satisfied indicators necessarily warrant an investigation.

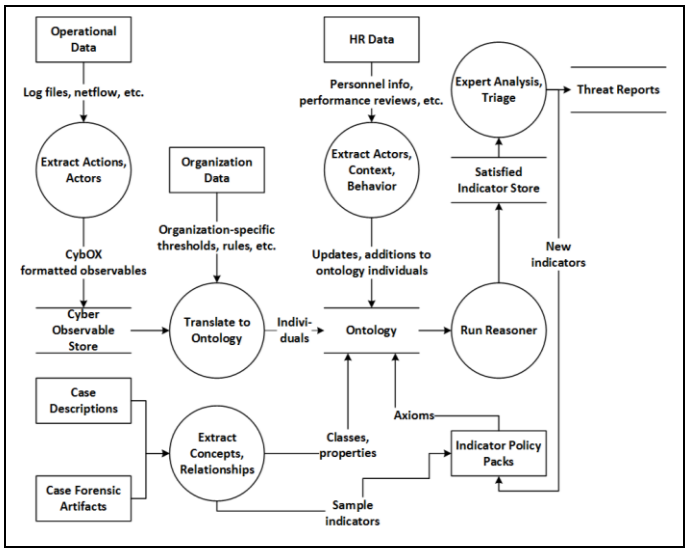

Fig. 5. Data flow diagram for automated indicator instance extraction framework

The evaluation of specific instances of indicators requires expert analysis and investigation to remove false positives, assess severity of the satisfied indicator, and perform set and temporal analysis on the satisfied indicators. The framework can support a workflow-based analysis and incident escalation process. Specific implementations of the framework are expected to grow and change as the organization, its insider threat program, and the larger insider threat community and domain all do the same. The activities associated with the operations and maintenance of this framework include

- Identifying new candidate indicators during the analysis of satisfied indicators
- Adding new indicators to the ontology as updates or additions to indicator policy packs
- Re-running the semantic reasoner as new ontology individuals are created and new indicators are added
- Adding automated ingest support for new operational data sources
- Extending the human resources data ingest process to include new data sources
- Updating the configuration for the automated ontology individual extractor as organizational policies change and new insights are gained

In addition to the activities mentioned above, the ontology itself will grow and change over time. The drivers for ontology changes will be the addition of new concepts and relationships based on analysis of new cases involving malicious insider activity, as well as feedback from the organizations that are using the ontology. Finally, indicator policy packs can be safely shared with other organizations as a means of identifying effective industry specific and domain-wide detection strategies and patterns.

VI. Conclusion

With the initial development of our ontology, we have created a bridge between natural language descriptions of potential indicators of malicious insider activity in case data and the operational data that contains the technical and

behavioral observables associated with malicious insider activity. We have provided a mechanism that allows sensitive information to be abstracted away while maintaining enough descriptive ability to effectively communicate actions and behaviors of interest across organizations. By introducing the application of our ontology as an analysis hub that combines operational and human resources data, we have laid the foundation for more effective fusion of these traditionally disparate data sources.

VII. Future Work

As we continue the development of our ontology, we will perform the following activities in future work:

- Provide enhanced support for behavioral components of potential indicators of malicious insider activity
- Collaborate with other organizations to improve the expression of insider threat indicators using the ontology
- Add support for additional indicator policy packs
- Mature the proof of concept automated indicator instance extractor and provide customization options for additional data sources and organization configurations
- Assess the feasibility of automating the creation of ontology individuals based on human resources data
- Evaluate formal ontology validation methods and apply them to our ontology

Acknowledgement

The authors gratefully acknowledge support for this work from the Defense Advanced Research Projects Agency (DARPA) and the Federal Bureau of Investigation. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Approved for Public Release, Distribution Unlimited.

References

[1] D. M. Cappelli, A. P. Moore, and R. F. Trzeciak, *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud)*: Pearson Education, 2012.

[2] U.S. GOVERNMENT, "Executive Order 13587-Structural Reforms To Improve the Security of Classified Networks and the Responsible Sharing and Safeguarding of Classified Information," 2011.

[3] B. Obama, "National Insider Threat Policy and Minimum Standards for Executive Branch Insider Threat Programs," T. W. House, Ed., ed: Office of the Press Secretary, 2012, p. 1.

[4] F. o. A. Scientists, "National Insider Threat Policy and Minimum Standards for Executive Branch Insider Threat Programs (Minimum Standards)," T. W. Hourse, Ed., ed. www.fas.org: Federation of American Scientists, 2012.

[5] M. West, *Developing high quality data models*: Elsevier, 2011.

[6] F. Intelligence and National Security Alliance (INSA) in partnership with DHS, and ODNI. (2014). Insider Threat Resource Directory. Available: http://www.insaonline.org/insiderthreat

[7] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, et al., "Gene Ontology: tool for the unification of biology," *Nature genetics*, vol. 25, pp. 25-29, 2000.

[8] S. Schulze-Kremer, "Adding semantics to genome databases: towards an ontology for molecular biology," in Ismb, 1997, p. 5.

[9] M. Grüninger and M. S. Fox, "The role of competency questions in enterprise engineering," in Benchmarking—Theory and Practice, ed: Springer, 1995, pp. 22-31.

[10] A. Gangemi, "Ontology design patterns for semantic web content," in The Semantic Web–ISWC 2005, ed: Springer, 2005, pp. 262-276.

[11] R. R. Starr and J. M. P. de Oliveira, "Conceptual maps as the first step in an ontology construction method," in Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2010 14th IEEE International, 2010, pp. 199-206.

[12] K. Žubrinic, "Automatic creation of a concept map."

[13] J. J. Villalon and R. A. Calvo, "Concept Map Mining: A definition and a framework for its evaluation," in Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on, 2008, pp. 357-360.

[14] S. Barnum, "Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™)," *MITRE Corporation, July*, 2012.

[15] The MITRE Corporation. (2014). *Cyber Observable eXpression*. Available: http://cybox.mitre.org/language/version2.1/

[16] R. Lee, "SANS Digital Forensics and Incident Response Poster Released," in Blog: SANS Digital Forensics and Incident Response Blog vol. 2014, S. D. Faculty, Ed., ed. SANS: SANS, 2012.

[17] *schema.org*. Available: http://schema.org

[18] J.-b. Gao, B.-w. Zhang, X.-h. Chen, and Z. Luo, "Ontology-based model of network and computer attacks for security assessment," *Journal of Shanghai Jiaotong University (Science)*, vol. 18, pp. 554-562, 2013/10/01 2013.

[19] S. Fenz and A. Ekelhart, "Formalizing information security knowledge," in Proceedings of the 4th international Symposium on information, Computer, and Communications Security, 2009, pp. 183-194.

[20] L. Obrst, P. Chase, and R. Markeloff, "Developing an ontology of the cyber security domain," Proceedings of Semantic Technologies for Intelligence, Defense, and Security (STIDS), pp. 49-56, 2012.

[21] S. E. Parkin, A. van Moorsel, and R. Coles, "An information security ontology incorporating human-behavioural implications," in Proceedings of the 2nd International Conference on Security of Information and Networks, 2009, pp. 46-55.

[22] M. Poveda-Villalón, M. C. Suárez-Figueroa, and A. Gómez-Pérez, "Validating ontologies with oops!," in *Knowledge Engineering and Knowledge Management*, ed: Springer, 2012, pp. 267-281.

[23] Aldo Gangemi. (2010). *Submissions:Sequence*. Available: http://ontologydesignpatterns.org/wiki/Submissions:Sequence

[24] G. Antoniou and F. Van Harmelen, "Web ontology language: Owl," in Handbook on ontologies, ed: Springer, 2004, pp. 67-92