Using Semantic Web Technologies to Develop Intrinsically Resilient Energy Control Systems

Frederick Sheldon and Daniel Fetzer
Oak Ridge National Laboratory
Oak Ridge, TN 37831, U.S.A.
{sheldonft, fetzerdt}@ornl.gov

Jingshan Huang
University of South Alabama
Mobile, AL 36688, U.S.A.
huang@southalabama.edu

Jiangbo Dang and Dong Wei
Siemens Corporation, Corporate Research and Technology
Princeton, NJ 08540, U.S.A.
{jiangbo.dang, dong.w}@siemens.com

David Manz
Pacific Northwest National Laboratory
Richland, WA 99354, U.S.A.
david.manz@pnnl.gov

Thomas Morris
Mississippi State University
Mississippi State, MS 39762, U.S.A.
morris@ece.msstate.edu

Jonathan Kirsch and Stuart Goose
Siemens Corporation, Corporate Research and Technology
Berkeley, CA 94704, U.S.A.
{jonathan.kirsch, stuart.goose}@siemens.com

***Abstract*—To preserve critical energy control functions while under attack, it is necessary to perform comprehensive analysis on root causes and impacts of cyber intrusions without sacrificing the availability of energy delivery. We propose to design an intrinsically resilient energy control system where we extensively utilize Semantic Web technologies, which play critical roles in knowledge representation and acquisition. While our ultimate goal is to ensure availability/resiliency of energy delivery functions and the capability to assess root causes and impacts of cyber intrusions, the focus of this paper is to demonstrate a proof of concept of how Semantic Web technologies can significantly contribute to resilient energy control systems.**

***Index Terms*—cybersecurity, energy control system, ontology, knowledge base, semantic annotation, data integration.**

I. Introduction

Our energy infrastructure depends on energy delivery systems comprised of complex and geographically dispersed network architectures with vast numbers of interconnected components. These systems provide critical functions to provide information and automated control over a large, complex network of processes that collectively ensure reliable and safe production and distribution of energy. The energy utilities are modernizing these vast networks with millions of smart meters, high speed sensors, advanced control systems, and a supporting communications infrastructure. This additional complexity brings benefits, but also increases the risks of cyber attacks that could potentially disrupt our energy delivery. These systems must maintain high availability and reliability even when under attack. After a security incident has been detected, the incident response team needs the ability to investigate and determine the root cause, attack methods, consequences, affected assets, impacted stakeholders, and other information in order to inform an effective response. The response team needs this information in the short term in order to contain or eradicate the attack, recover compromised equipment, and restore normal operation. The team also needs to determine counter-measures to prevent recurrence and possibly collect evidence to legally prosecute the offenders. This analysis and response must be done without interrupting the availability of the energy delivery systems.

To address the aforementioned challenges, this paper presents the design and architecture of *InTRECS*, an InTrinsically Resilient Energy Control System. The ultimate goal of InTRECS is to provide tools and technologies to ensure the availability/resiliency of energy delivery functions, along with the capability to assess root causes and impacts of cyber intrusions. To meet these goals, InTRECS extensively applies Semantic Web technologies, including cybersecurity domain ontologies, a comprehensive knowledge base, and semantic data annotation & integration techniques. Semantic Web technologies are built upon ontologies, which are formal, declarative knowledge models and have been shown to play critical roles in knowledge representation and acquisition.

In this paper, we argue that applying Semantic Web technologies in InTRECS affords several benefits compared to typical approaches that utilize relational databases:

- While relational databases focus on *syntactic* representation of data and lack the ability to explicitly encode semantics, Semantic Web technologies support rich *semantic* encoding, which is critical in automated knowledge acquisition.
- Powerful tools exist for capturing and managing ontological knowledge, including an abundance of reasoning tools readily supplied for ontological models, making it much more convenient to query, manipulate, and reason over available data sets. As a result, semantics-based queries, instead of SQL queries, are made possible.
- Advances in an energy delivery system (EDS) require changes to be made regularly regarding underlying data models. In addition, more often than not, it is preferable to represent data at different levels and/or with different abstractions. There are no straightforward methods for performing such updates if relational models are adopted.
- Semantic Web technologies better enable EDS researchers to append additional data into repositories in a more

flexible and efficient manner. The formal semantics encoded in ontologies makes it possible to reuse data in unplanned and unforeseen ways, especially when data users are not data producers, which is now very common.

While our ultimate goal is to ensure availability/resiliency of energy delivery functions and the capability to assess root causes and impacts of cyber intrusions, the focus of this paper is to demonstrate a proof of concept of how Semantic Web technologies can significantly contribute to resilient energy control systems. The rest of the paper is organized as follows. Section II gives a brief review on related research in ontologies and semantic annotation & integration, respectively. Section III describes the overall architecture of InTRECS, followed by methodology details for developing domain ontologies & knowledge base and performing data annotation & integration. Section IV demonstrates our preliminary experimental results. Finally, Section V concludes with future research directions.

II. Related Work

A. Ontologies in Energy Delivery Control and Cybersecurity

Energy delivery control systems comprise complex network architectures that may contain hundreds of specialized cyber components and may extend across wide geographical regions. Cyber attack investigation involves examining large volumes of data from heterogeneous sources. Researchers are facing the challenge of how to maintain the integrity of data derived from diverse sources across distributed geographic areas ([1-7]). These research efforts have resulted in various ad-hoc proprietary formats for storing and analyzing data and maintaining respective metadata. Different parties are likely to adopt different formats according to specific needs. Therefore, the seamless communication among different parties, along with the knowledge sharing and reuse that follow, become a non-trivial problem. Turnitsa and Tolk [8] discussed in depth multi-resolution, multi-scope, and multi-structure challenges during data exchange between different models.

Semantic Web technologies that are based on domain ontologies can render tremendous help. Ontologies are declarative knowledge models, defining essential characteristics and relationships for specific domains of interest. As a semantic foundation, ontologies greatly help domain experts to formally define domain knowledge in terms of *data semantics* (intended meanings) rather than *data syntax* (forms in which data are represented). Reasons for developing ontologies include, but not limited to: (i) to share domain information among people and software; (ii) to enable reuse of domain knowledge; (iii) to analyze domain knowledge and make it more explicit; and (iv) to separate domain knowledge from its implementation. There exist some domain ontologies in cybersecurity and related areas, e.g., Intrusion Detection System Ontology [1], Network Security Ontology [2], Process Control Ontology [4], INSPIRE Ontology [5], and GE SADL Host Defense Ontology [7]. These ontologies provide metadata and standard terminologies in respective domains.

B. Semantic Data Annotation & Integration

Semantic data annotation & integration can bring critical impacts and benefits to data analysis and management. Semantic annotation (tagging) systems can be divided into manual, semi-automatic, and automatic ones [9]. In manual tagging systems (Sema-Link [10] for example), users employ controlled vocabularies from some ontology to tag documents. Such a manual process is time-consuming and requires deep domain expertise, in addition to the inconsistency issue. Semi-automatic tagging systems improve manual tagging systems by automatically parsing documents and recommending potential tags. Human annotators only need to select tags from candidates suggested by the system. Automatic semantic tagging systems offer further improvement by parsing and tagging documents with ontological concepts and instances in a fully automatic way. Zemanta [11] is such an example. By suggesting contents from various sources, such as Wikipedia, YouTube Flickr, and Facebook, Zemanta disambiguates terms and maps them to the Common Tag Ontology [12]. Dang et al. have developed one of the largest comprehensive, domain-independent ontological knowledge base, UNIpedia+ [13], which covers around 11 million named English entities. Based on UNIpedia+, they further developed an automatic tagging system [14] to produce semantically linked tags for given data. The information system architecture in the Los Angeles Smart Grid project [15] enabled analytical tools and algorithms to forecast energy load and identify load curtailment response through semantically meaningful data.

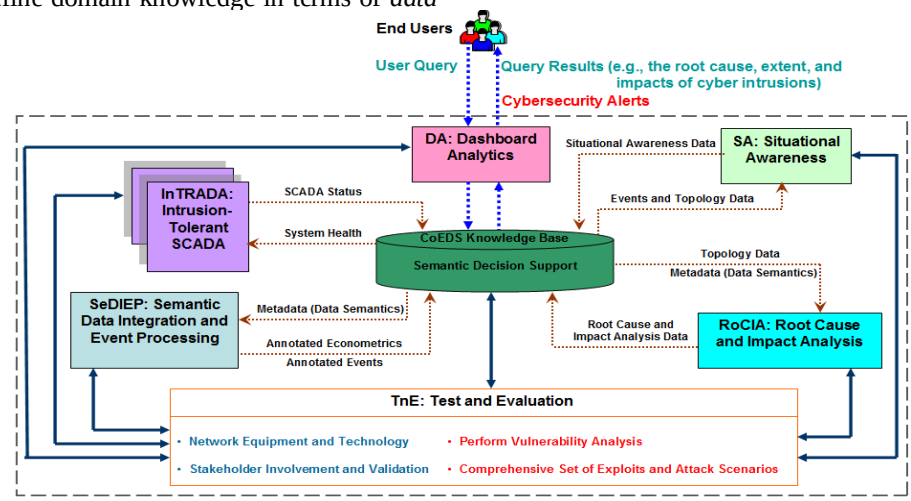

Fig. 1. Overall architecture of InTRECS system.

III. METHODOLOGY

A. InTRECS Overall Architecture

Figure 1 illustrates the overall architecture of InTRECS, which is decomposed into six subsystems.

- *Intrusion-Tolerant SCADA (InTRADA)*
  We will develop a survivable SCADA system based on intrusion-tolerant replication [16]. InTRADA will be capable of guaranteeing correct operations and excellent performance even when part of the system has been compromised and is under the control of an intelligent attacker.
- *Cybersecurity Ontologies and Knowledge Base for Energy Delivery Systems (CoEDS)*
  CoEDS knowledge base (KB) contains domain ontologies, a resource description framework (RDF) repository, a SPARQL RDF query engine, and an inference engine. The KB will provide end users with a unified and consistent data layer for analyzing data at the semantic level.
- *Semantic Data Integration and Processing (SeDIEP)*
  Our focus is to develop an automatic semantic data annotation & integration engine for tagging data sources based on the metadata defined in CoEDS ontologies. An event-processing engine will handle dynamic events and generate security alerts.
- *Root Cause and Impact Analysis (RoCIA)*
  RoCIA provides the basis to detect cyber incidents and investigate the root cause, attack methods, consequences, affected assets, impacted stakeholders, attackers' identity, and other metrics to inform an effective response. RoCIA will leverage the Cyber Security Econometrics System (CSES) and the inference and query engines provided within CoEDS KB to assist EDS stakeholders in evaluating cybersecurity investments and to provide an economic impact assessment of on-going cyber intrusions.
- *Dashboard Analytics and Situation Awareness (DaSA)*
  Dashboard analytics includes a user graphical user interface (GUI) to support interactions between end users and InTRECS. Situational awareness will be performed for end users. We will also support reasoning through the inference engine in CoEDS.
- *Test and Evaluation (TnE)*
  Implemented modules will automatically configure the test suite environment to the appropriate start state for the test case. A portal will provide the information and documentation and will execute the test case. We will also develop a test suite in an end-user setting, including a set of denial of service (DOS), reconnaissance, and network packet integrity exploits targeting SCADA, remote terminal unit (RTU), and network architecture vulnerabilities.

InTRECS will be constantly active to intrinsically provide resiliency, i.e., correct operations and excellent performance. At the same time, a DaSA GUI will guide end users to generate queries out of data derived from diverse sources. Query results, e.g., the root cause, extent, and impacts of the cyber intrusion, can then be provided back to end users. InTRECS will also push security alerts up to end users. Both query results and alerts are regarded as semantic decision support to end users because they extensively utilize Semantic Web technologies, namely, domain ontologies, RDF triples resulting from semantic annotation, and inferences & analysis performed at the semantic level.

B. CoEDS Domain Ontologies and Knowledge Base

There are four components in CoEDS KB: (i) CoEDS domain ontologies, (ii) an RDF repository, (iii) a SPARQL RDF query engine, and (iv) an inference engine. Through automatic data integration and logic reasoning, CoEDS KB will be able to provide a unified and consistent data layer for analyzing data *at the semantic level*. It will thus assist end users to effectively obtain real-time decision support, so that they can (i) obtain health status updates of SCADA replicas, (ii) analyze and better understand the root cause, extent, and impacts of an attack, (iii) acquire situational awareness, and (iv) recommend courses of action.

1) *Interaction between CoEDS and other InTRECS subsystems:* CoEDS KB actively exchanges information with other subsystems of InTRECS on a regular basis.

- InTRADA receives system health and status information from CoEDS KB, and incorporates such knowledge to enhance its fault-detection algorithms. This will enable InTRADA to more rapidly reconfigure itself in the event of a cyber attack by helping it distinguish between performance faults caused by a malicious application and by more benign issues such as transitory network problems. InTRADA sends to CoEDS KB status updates regarding the health of the replicas, hence providing data for future cyber attack analysis.
- SeDIEP obtains the data semantics, i.e., ontological metadata, from CoEDS KB and utilizes such metadata during the automatic semantic annotation. Annotated data, including cybersecurity econometrics, dynamic events, etc., are stored back into CoEDS KB to construct and continuously update the central data repository in the KB.
- CoEDS KB provides RoCIA with topology data as well as the data semantics essential for performing root cause and impact analysis. RoCIA supplies CoEDS KB with root cause and impact analysis data, including attack signatures, attack locations, exploits, consequences, countermeasures, model parameters, network components, security requirements, threats, vulnerabilities, and stakeholders.
- CoEDS KB furnishes DaSA with dynamic events and electric grid components and topology data, both of which are in an annotated form. DaSA sends back situational awareness data to CoEDS KB. In addition, the KB also provides the Correlation Layers for Information Query and Exploration (CLIQUE) and Traffic Circle, two visual analytics tools in DaSA, with interoperability for behavior model-based anomaly detection.

2) *Motivation for developing CoEDS ontologies:* Among existing ontologies in cybersecurity and related areas (mentioned in Section II), there is *not a single one that is comprehensive enough* to cover a complete set of concepts and relationships for the purpose of this research. In particular, with regard to the fields of SCADA status, root cause analysis, situational awareness, electric grid components and topology, cybersecurity econometrics, cost benefit analysis, and complex event processing, all aforementioned existing ontologies are missing some necessary concepts within these critical fields. Even in the case that a specific concept of our interest is contained in some existing ontology, more often than not, the semantics defined in such an ontology need to be extended and customized before this concept can be utilized within InTRECS system. In brief, Energy Control Systems (ECS) end users lack a comprehensive, customized conceptual model, which prevents the energy sector from leveraging enhanced knowledge acquisition processes brought by Semantic Web technologies. Such a situation motivates us to develop CoEDS domain ontologies.

3) *Ontology development principles:* We have observed seven practices suggested by Smith et al. [17]: the ontology should (i) be freely available; (ii) be expressed using a standard language or syntax; (iii) provide tracking and documentation for successive versions; (iv) be orthogonal to existing ontologies; (v) include natural language specifications of all concepts; (vi) be developed collaboratively; and (vii) be used by multiple researchers. In particular, we propose a *decomposition* methodology as the strategy for coming up with orthogonal ontologies. Our methodology is similar to those used in the database normalization theory, third normal form (3NF) for example. We first began with concepts from possibly many sub-domains in one large set, followed by the identification of dependencies or overlaps among these concepts, and we finally proceeded to decompose all concepts based on their identified dependencies. Our preliminary design is to develop seven sub-ontologies in CoEDS: SCADA status, root cause & impact, situational awareness, grid component & topology, cybersecurity econometrics, cost benefit, and complex event processing. Consequently, we achieved the orthogonality feature, i.e., the non-overlapping feature, for CoEDS domain ontologies.

4) *Knowledge-driven ontology development procedure:* The ontology development was not from scratch. Instead, to (i) take advantage of the knowledge already contained in existing ontologies and (ii) reduce the possibility of redundant efforts, we have reused, extended, and customized a set of well-established concepts from existing domain ontologies. In addition, popular upper ontologies, e.g., the Basic Formal Ontology (BFO), was imported into our ontologies. The ontology development was driven by domain knowledge and decomposed into five stages, as suggested by Uschold and Gruninger [18]: (i) specification of content; (ii) informal documentation of concept definitions (by domain experts); (iii) logic-based formalization of concepts and relationships between concepts; (iv) implementation of the ontology in a computer language; and (v) evaluation of the ontology, including the internal consistency and the ability to answer logical queries. As illustrated in Figure 2, these five stages are essentially ongoing and iterative because end users' needs will change as their understanding of the domain evolves. In this iterative, knowledge-driven approach, both ontology engineers and domain experts have been involved, working together to capture domain knowledge, develop a conceptualization, and implement the conceptual model. The ontology construction process has taken place over a number of iterations, involving a series of interviews, evaluation strategies, and refinements. Standard revision-control procedures have been utilized.

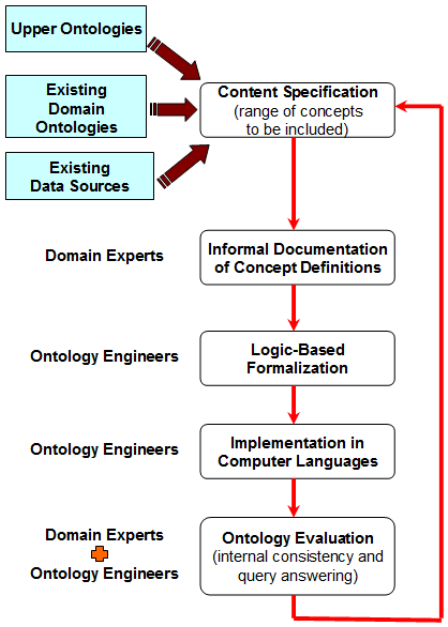

Fig. 2. Knowledge-driven, iterative ontology development.

5) *Ontology format and development tool:* There are different formats and languages for describing ontologies, all of which are popular and based on different logics: Web Ontology Language (OWL) [19], Open Biological and Biomedical Ontologies (OBO) [20], Knowledge Interchange Format (KIF) [21], and Open Knowledge Base Connectivity (OKBC) [22]. We have chosen the OWL format recommended by the World Wide Web Consortium (W3C). OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans. As a result, OWL facilitates greater machine interpretability of Web contents. We have chosen Protégé, an open-source ontology editor developed by

Stanford [23], as our development tool over other available tools such as CmapTools and OntoEdit.

6) CoEDS KB components – RDF Repository, Query Engine, and Inference Engine: Based on the formal knowledge defined in CoEDS ontologies, heterogeneous data sources can be annotated and integrated into a central repository. Note that data sources to be integrated include structured, semi-structured, or unstructured data, the interoperability thus becomes an obstacle during knowledge discovery. We adopt RDF, a model for data interchange recommended by the W3C, to handle such a challenge. RDF specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. The generic structure of RDF allows structured, semi-structured, and unstructured data to be mixed, exposed, and shared across different applications, thus helping to handle the data interoperability challenge. Following automatic semantic data annotation (see Section III.C), RDF triples will be indexed and accumulated into a central repository. SPARQL Protocol and RDF Query Language (SPARQL) [24] is a query language recommended by W3C to retrieve and manipulate RDF data. End users of InTRECS system will be guided by a GUI to automatically generate RDF queries across semantically integrated sources. These queries will then be executed by a SPARQL-based query engine.

The RDF data repository and query answering are not enough for an effective and comprehensive knowledge acquisition. Suppose that some facts do not exist in any original data sources, they will thus not be stored in the RDF repository. But such information may be critical to end users. To obtain the ability to acquire previously implicit knowledge, we will incorporate an inference engine (a.k.a. logic reasoner). Compared with traditional relational database techniques, inference engines provide a more expressive method for querying and reasoning over available data sets. Thus, ontology-based (a.k.a. semantics-based) queries, instead of traditional SQL queries, are possible. Ontology-based queries improve traditional keyword-based queries in several ways. (i) Both *synonymous* terms (those having same meaning) and *polysemous* terms (those having different meanings) can be included to obtain more results that are relevant to the user query. (ii) Semantic relationships among terms often reveal extra clues hidden in disparate data sources. Such relationships can be explicitly discovered to further improve the quality of query answering. Consequently, we will be able to acquire hidden knowledge and information that was originally implicit and unclear, yet critical, to end users. With a logic reasoner, CoEDS repository will work as a comprehensive knowledge base.

7) *Sesame framework for RDF repository, SPARQL RDF query engine, and inference engine:* We have preliminarily chosen Sesame framework [25] to store and manage the RDF repository. Sesame is an open-source Java framework for the storage and querying of RDF data. The framework is fully extensible and configurable with respect to storage mechanisms, inferencers, RDF file formats, query result formats, and query languages. In addition, Sesame offers a JBDC-like user API, streamlined system APIs, and a RESTful HTTP interface supporting the SPARQL protocol for RDF. Moreover, Sesame contains a built-in inference engine, and various reasoning tasks, e.g., subsumption and contradiction reasoning, can be performed.

C. Semantic Data Annotation and Event Processing

According to the formal domain knowledge, including a global metadata model, defined in CoEDS, heterogeneous data sources can be annotated and seamlessly integrated into a central RDF data repository, which will serve as a unified and consistent data layer for data analytics applications.

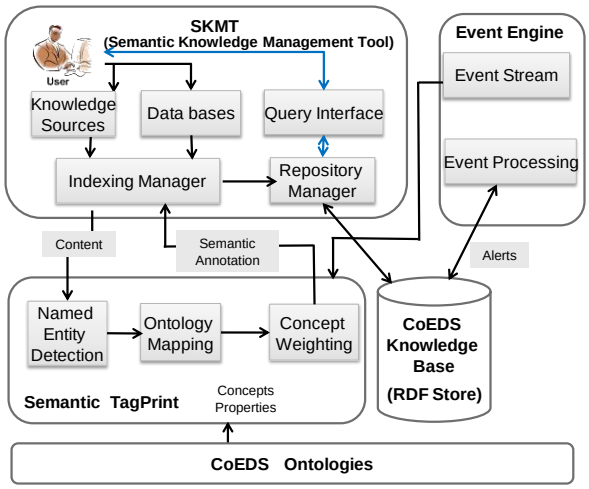

Fig. 3. Semantic data annotation and event processing (SeDIEP).

1) System overview: Semantic data annotation and event processing (SeDIEP) subsystem manages various data sources and automatically annotates and integrates data at semantic level. As shown in Figure 3, there are three major components in the subsystem: (i) Semantic TagPrint, (ii) Semantic Knowledge Management Tool (SKMT), and (iii) Event Engine. Semantic TagPrint is an automatic semantic tagging engine that annotates structured data and free text using ontological entities from CoEDS ontologies. SKMT manages heterogeneous data sources for semantic annotation and integration. Event engine feeds the semantic tagging engine with dynamic events. It also generates alerts with the support from CoEDS through modified RDF queries and the semantic reasoning.

Heterogeneous data sources will be annotated and seamlessly integrated into a central RDF data repository based on CoEDS ontologies. This data repository will serve as a unified and consistent data layer for further analyzing data at the semantic level. Our core technologies can substantially reduce design-to-execution time for application domains of data integration, visualization, and analysis.

- Meaningful data. Our system will annotate terms in text with their corresponding concepts in CoEDS ontologies by finding their meanings and analyzing their context.
- Scalability. Indexed data are stored and managed in a repository. Collected and initially processed data can be incrementally analyzed and indexed.
- Easy integration. Various data sources can be seamlessly integrated along with their semantic indexes.

2) *Deep annotation and integration:* Data sources to be integrated contain structured, semi-structured, or unstructured data. As discussed in the previous section, we adopt RDF to handle the data interoperability challenge. Semantic data annotation is the process of tagging source files with metadata predefined in ontologies such as names, entities, attributes, definitions, and descriptions. Herein, we use terms of "semantic annotation" and "semantic tagging" interchangeably. The annotation provides extra information contained in metadata to existing pieces of data. Metadata are usually from a set of ontological entities (including concepts and instances of concepts) predefined in ontologies. For unstructured data such as free text, we will use a tagging engine to align them with ontological entities and generate semantic annotations. For structured data including database data, the annotation will take two successive steps: (i) first we will annotate data source schemas by aligning their metadata with ontological entities; (ii) according to annotated schemas we will then transform original data instances into RDF triples. We refer to such annotation as "deep" annotation – this term was coined by Goble, C. in the Semantic Web Workshop of WWW 02. It is necessary to annotate more than just data source schemas because there are situations where the opposite "shallow" annotation (i.e., annotation on schemas alone) cannot provide users with the desired knowledge. Following semantic data annotation, RDF triples will be indexed and accumulated into a central repository.

3) *Unified view over original data sources and cost-efficient analysis:* All semantic tags will be generated from a global metadata model, i.e., CoEDS ontologies, our tool thus provides a unified view over original data sources at the semantic level. As discussed before, our RDF query and reasoning engines will provide users with more meaningful and relevant information from semantically annotated and integrated data sources. In addition, semantic relationships among tags provide us with additional clues and will further improve the quality of retrieved results. Given a set of candidate results to be returned to users, we will calculate the semantic similarity between each result and the user query using semantic features such as (i) *hypernym*, which defines the *superClassOf* relationship and (ii) *holonym*, which defines the *partOf* relationship. We will then rank these results by their respective semantic similarities. Consequently, users can be presented with more relevant query results.

4) *Semantic event processing:* Dynamic events will be fed to our Semantic Tag Print, which will annotate these events with semantic tags. Then events are represented as RDF triples, accompanied with event attributes such as timestamps and probabilities. With the support from CoEDS, SeDIEP will transform these tagged events into SPARQL queries. We will perform event filtering, correlation, and aggregation or abstraction using semantic matching, rules, and similarity evaluations. Moreover, we will detect event patterns on event streams with temporal semantic rules. As a result, high-risk vulnerabilities and threats can be predicted, and security alerts will then be automatically generated and rendered to users when facing potential cyber intrusions.

5) *Core Components in SeDIEP:* Figure 3 shows three major components in SeDIEP to semantically integrate various data sources and event streams.

a) Component one: Semantic TagPrint is an automatic semantic tagging engine that annotates structured data and free text using ontological entities. Three modules were designed for this component.

- Named Entity Detection: This module extracts named entities, noun phrases in general, from the input text. We adopt Stanford Parser [26] to detect and tokenize sentences, and assign Part-of-Speech (PoS) tags to tokens. Entity names will be extracted based on PoS tags.
- Ontology Mapping: This module maps extracted entity names to CoEDS concepts and instances with two steps: Phrase mapping and Sense mapping. Phrase mapping will match the noun phrase of an entity name to a predefined concept or instance. Sense mapping will utilize a linear-time lexical chain algorithm to disambiguate terms that have several senses defined in ontologies.
- Ontology Weighting: This module utilizes statistical and ontological features of concepts to weigh semantic tags. We then annotate the input text using the semantics with higher weights.

b) Component two: SKMT collects original text and sends annotation results to Repository Manager, whose main role is to manage RDF repository (store) and to communicate with Query Interface. These components altogether provide a unified view over original data sources at the semantic level. Users will be guided by a GUI to automatically generate RDF queries across semantically integrated data sources. These queries will then be executed by a SPARQL-based RDF query engine. As discussed earlier in this subsection, we can calculate the semantic similarity between each candidate query result and the user query using semantic features such as *hypernym* and *holonym*. These query results can then be ranked by their respective semantic similarities. Consequently, we are able to render users more accurate and desired query results.

c) Component three: Event Engine annotates dynamic events and stores them as RDF triples. It will then generate SPARQL queries and perform event filtering, correlation, and aggregation or abstraction with the semantics defined in CoEDS ontologies.

IV. Preliminary Experimental Results

In this ongoing research, we have developed a preliminary version of CoEDS domain ontologies and knowledge base to demonstrate a proof of concept of how Semantic Web technologies can significantly contribute to resilient energy control systems. We also exported instances into an RDF data repository within the Sesame framework.

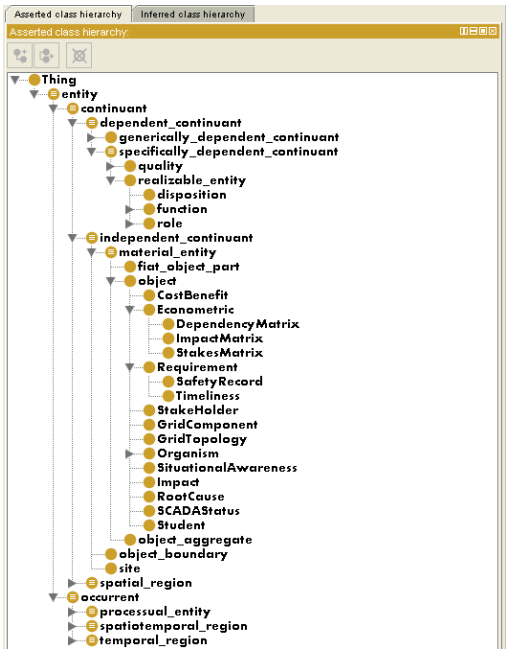

Fig. 4. Protégé GUI screen shot exhibiting some CoEDS concepts.

A. CoEDS Ontologies

As discussed earlier in Section III.B, we have developed seven sub-ontologies in CoEDS: *SCADA Status Ontology*, *Root Cause & Impact Ontology*, *Situational Awareness Ontology*, *Grid Component & Topology Ontology*, *Cybersecurity Econometrics Ontology*, *Cost Benefit Ontology*, and *Complex Event Processing Ontology*. The purpose of such a decomposition strategy is to achieve the orthogonality feature, i.e., the non-overlapping feature among different CoEDS sub-ontologies. After individual sub-ontologies were developed, we then imported them into CoEDS. If future modifications are needed for any sub-ontology, the changed schema information will be automatically integrated into CoEDS ontologies. Figure 4 demonstrates a screen shot from Protégé GUI, which exhibits a portion of CoEDS concepts. Note that the well-defined, general-purpose structure from the Basic Formal Ontology (BFO), a popular upper ontology across different disciplines and research areas, was preserved in the ontology schema. Statistic information for all seven sub-ontologies is summarized in Table I. In total, CoEDS ontologies contain 269 concepts, 232 object properties, and 110 data properties.

TABLE I. Statistics for CoEDS Ontologies

Sub-Ontology	**Statistic Information**		
	Total Number of Concepts	***Total Number of Object Properties***	***Total Number of Data Properties***
SCADA Status Ontology	35	23	12
Root Cause & Impact Ontology	37	21	9
Situational Awareness Ontology	39	27	15
Grid Component & Topology Ontology	51	39	17
Cybersecurity Econometrics Ontology	38	25	20
Cost Benefit Ontology	33	19	18
Complex Event Processing Ontology	36	28	19

B. CoEDS Knowledge Base

The current CoEDS KB contains a total of 1,223 facts (a.k.a. axioms in Protégé). Details can be found in Table II.

TABLE II. Statistics for CoEDS Knowledge Base Axioms

Axiom Category	**Statistic Information**
Class Axioms	460
Subclass Axioms	268
Equivalent Class Axioms	57
Disjoint Class Axioms	135
Object Property Axioms	217
Data Property Axioms	108
Individual Axioms	236
Annotation Axioms	202

C. Sesame Framework to Manage Data Repository

Within the Sesame framework we exported all ontological instances into an RDF data repository for future storage and management. Figure 5 is a screen shot from Sesame GUI, where the seven sub-ontologies and the overall CoEDS ontologies were clearly demonstrated. Being an open-source Java framework, Sesame framework can be readily extended and configured for the storage and querying of RDF data. Moreover, a JBDC-like user API, streamlined system APIs, and a RESTful HTTP interface are offered in Sesame as well.

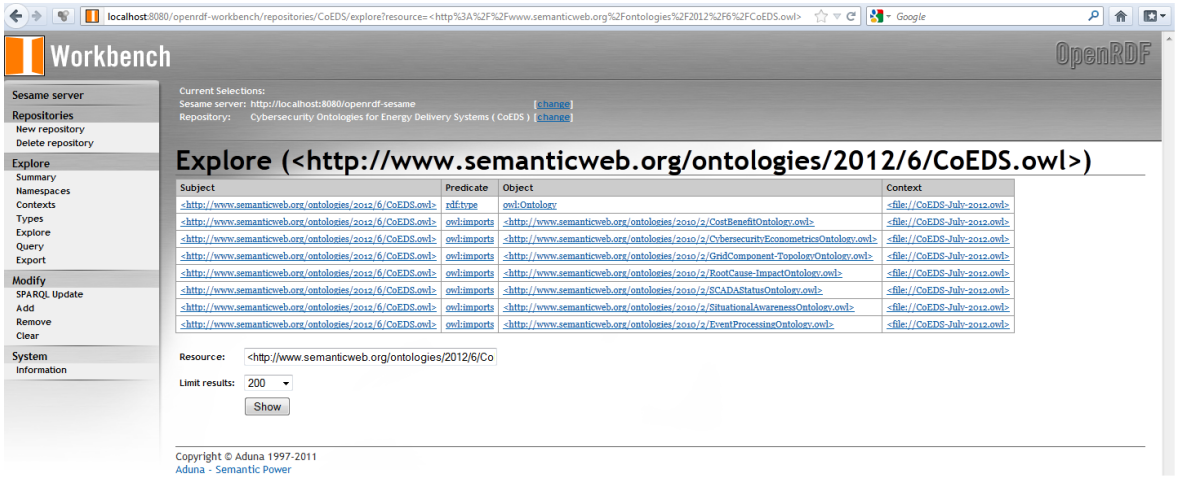

Fig. 5. Screen shot from Sesame repository management.

V. Conclusion

To preserve critical energy control functions while under attack, it is necessary to perform comprehensive analysis on the root cause, extent, and impacts of cyber intrusions without sacrificing the availability of energy delivery. We proposed to develop InTRECS, an intrinsically resilient energy control system, to address these challenges. Semantic Web technologies, which play critical roles in knowledge representation and acquisition, have been extensively adopted in our system. The focus of this ongoing research is to demonstrate a proof of concept of how Semantic Web technologies can significantly contribute to resilient energy control systems. We justified the research motivation, described our methodology in detail, and exhibited preliminary experimental results. Future research directions include, but are not limited to, (i) continue CoEDS ontology development towards delivering a highly stable and more usable version; (ii) incorporate query and inference engines into the knowledge base for end users to better analyze root causes and impacts of cyber intrusions; and (iii) implement SeDIEP subsystem.

Acknowledgment

This research was partially supported through the U.S. Department of Energy (DOE) Higher Education Research Experiences (HERE) program for Faculty at the Oak Ridge National Laboratory, Oak Ridge, Tennessee, sponsored by the U.S. Department of Homeland Security (DHS).

References

[1] J. Undercoffer, A. Joshi, and J. Pinkston, "Modeling Computer Attacks: An Ontology for Intrusion Detection," *RAID 2003, LNCS 2820*, pp. 113-135, 2003, Springer-Verlag Berlin Heidleberg, 2003.

[2] A. Simmonds, P. Sandilands, and L. Ekert, "An Ontology for Network Security Attacks," *Proc. the 2nd Asian Applied Computing Conference (AACC-04)*, LNCS 3285, pp. 317-323, 2004.

[3] W. Wang and T. Daniels, "A Graph Based Approach toward Network Forensic Analysis," *ACM Transactions on Information and Systems Security*, Vol. 12, No. 1, Article 4, Pub. Date: Oct. 2008.

[4] J. Hieb, J. Graham, and J. Guan, "An Ontology for Identifying Cyber Intrusion Induced Faults in Process Control Systems," *Critical Infrastructure Protection III*, IFIP AICT 311, pp. 125-138, 2009.

[5] G. Isaza, A. Castillo, M. Lopez, L. Casillo, and M. Lopez, "Intrusion Correlation Using Ontologies and Multi-agent Systems," *Proc. 4th International Conference on Information Security and Assurance (ISA-10)*, pp. 355-361, Miyazaki, Japan, June 23-25, 2010.

[6] M. Choras, R. Kozik, A. Flizikowski, and W. Holubowicz, "Ontology Applied in Decision Support System for Critical Infrastructures Protection," *IEA/AIE2010*, LNAI, pp. 671-680, 2010.

[7] B. Barnett, A. Crapo, and P. O'Neil, "Experiences in Using Semantic Reasoners to Evaluate Security of Cyber Physical Systems," *General Electric Internal Report GridSec*, 2012.

[8] C. Turnitsa and A. Tolk, "Knowledge Representation and the Dimensions of a Multi-Model Relationship," *Proc. the 40th Conference on Winter Simulation (WSC-08)*, pp. 1148–56, 2008.

[9] L. Reeve and H. Han, "Semantic Annotation for Semantic Social Networks Using Community Resources," *AIS SIGSEMIS Bulletin*, vol. 2, pp. 52-56, 2005.

[10] S. Wiesener, W. Kowarschick, and R. Bayer, "SemaLink: An Approach for Semantic Browsing through Large Distributed Document Spaces," *Proc. the 3rd International Forum on Research and Technology Advances in Digital Libraries*, p. 86, 1996.

[11] Zemanta. http://www.zemanta.com/.

[12] Common Tag. http://www.commontag.org/.

[13] K. Murat, J. Dang, and S. Uskudarli, "UNIpedia: A Unified Ontological Knowledge Platform for Semantic Content Tagging and Search," *Proc. the 4th IEEE International Conference on Semantic Computing*, Pittsburg, PA, USA, 2010.

[14] K. Murat, J. Dang, and S. Uskudarli, "Semantic TagPrint: Indexing Content at Semantic Level," *Proc. the 4th IEEE International Conference on Semantic Computing*, Pittsburg, PA, USA, 2010.

[15] Y. Simmhan, Q. Zhou, and V.K. Prasanna,"Semantic Information Integration for Smart Grid Applications," *Chapter 19, Green IT: Technologies and Applications*, pp. 361–80, 2011.

[16] J. Kirsch, S. Goose, Y. Amir, and P. Skare, "Toward Survivable SCADA," *Proc. the Annual Cyber Security and Information Intelligence Research Workshop (CSIIRW-11)*, Oak Ridge, 2011.

[17] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. Goldberg, K. Eilbeck, A. Ireland, C. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S. Sansone, R. Scheuermann, N. Shah, P. Whetzel, and S. Lewis, "The OBO foundry: coordinated evolution of Ontologies to support biomedical data integration," *Nature Biotechnology*, 25(11):1251–1255, 2007.

[18] M. Uschold and M. Gruninger, "Ontologies: principles, methods, and applications," *Knowledge Engineering Review*, 11(2):93-155, 1996.

[19] OWL. http://www.w3.org/2004/OWL/.

[20] OBO. http://www.obofoundry.org/.

[21] KIF (Knowledge Interchange Format). http://logic.stanford.edu/kif/.

[22] OKBC. http://www.ai.sri.com/okbc/.

[23] Protégé. http://protege.stanford.edu/.

[24] SPARQL. http://www.w3.org/TR/rdf-sparql-query/.

[25] Sesame. http://www.openrdf.org/doc/sesame/.

[26] D. Klein and C.D. Manning, "Accurate Unlexicalized Parsing," *Proc. the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430, 2003.