PNNL-24871

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Federating Cyber and Physical Models for Event-Driven Situational Awareness

October 2015

EG Stephan
RA Pawlowski
S Sridhar
MJ Rice

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.
(8/2010)

PNNL-24871

Federating Cyber and Physical Models for Event-Driven Situational Awareness

EG Stephan
RA Pawlowski
S Sridhar
MJ Rice

October 2015

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Summary

This report provides details on our approach to data federation and how data interfaces between models will ultimately be used to build Unified Modeling Language profiles that can be more easily referenced and cataloged. This is the third technical report provided on this project. The two previous reports were focused on defining an example use cases in which federation of cyber-physical models are necessary. The report presents the interfaces between models for two use cases, cyber-physical contingency analysis and STEVE, or Security Threat EValuation Environment, our proposed suite of cyber-physical security evaluation applications that has the capability to identify cyber-attack scenarios in the power system Supervisory Control and Data Acquisition (SCADA) environment. The report concludes in a discussion of lessons learned and identifies next steps.

Acronyms

CA	Contingency Analysis
CCNA	Control Center Network Administrator
CIM	Common Information Model
CPCA	Cyber-Physical Contingency Analysis
CPS	Cyber-Physical System
CSO	Cyber System Operator
CVE	Common Vulnerabilities and Exposures
DDL	Data Definition Language
DMTF	Distributed Management Task Force
EMS	Energy Management System
FEDSEC	Federation of Utility Control Center Cybersecurity and Operations Data
IEC	International Electrotechnical Commission
NVD	National Vulnerability Database
OWL	Web Ontology Language
PC	Personal Computer
PGO	Power Grid Operator
PNNL	Pacific Northwest National Laboratory
RDFS	Resource Description Framework Schema
SCADA	Supervisory Control and Data Acquisition
SNA	Substation Network Administrator
SQL	Structured Query Language
STEVE	Security Threat EValuation Environment
STIX™	Structured Threat Information Expressions
TAXII	Trusted Automated eXchange of Indicator Information
TC	Technical Committee
TSNA	Target Substation Network Administrator
TTP	Tactics, Techniques, and Procedures
UML	Unified Modeling Language
USB	Universal Serial Bus
XMI	XML Metadata Interchange
XML	eXtensible Markup Language
XSD	XML Schema Definition

Contents

Figures

1.0 Introduction

The purpose of this paper is to describe a novel method to improve electric power system monitoring and control software application interoperability. This method employs the concept of federation, which is defined as the use of existing models that represent aspects of a system in specific domains (such as physical and cyber security domains) and building interfaces to link all of domain models. Federation seeks to build on existing bodies of work. Some examples include the Common Information Models (CIM) maintained by the International Electrotechnical Commission Technical Committee 57 (IEC TC 57) for the electric power industry. Another relevant model is the CIM maintained by the Distributed Management Task Force (DMTF); this CIM defines a representation of the managed elements in an Information Technology (IT) environment. The power system is an example of a cyber-physical system, where the cyber systems, consisting of computing infrastructure such as networks and devices, play a critical role in the operation of the underlying physical electricity delivery system. Measurements from remote field devices are relayed to control centers through computer networks, and the data is processed to determine suitable control actions. Control decisions are then relayed back to field devices. It has been observed that threat actors may be able to successfully compromise this cyber layer in order to impact power system operation. Therefore, future control center applications must be wary of potentially compromised measurements coming from field devices. In order to ensure the integrity of the field measurements, these applications could make use of compromise indicators from alternate sources of information such as cyber security. Thus, modern control applications may require access to data from sources that are not defined in the local information model. In such cases, software application interfaces will require integration of data objects from cross-domain data models. When incorporating or federating different domains, it is important to have subject matter experts work together, recognizing that not everyone has the same knowledge, responsibilities, focus, or skill set.

1.1 Model Federation Strategy

Initially, the federation of domain models can seem daunting since cyber and physical models contain literally thousands of classes and relationships between those classes. However, when considering that the motivation for model federation should only occur by use case requirements, federation can occur at a gradual level and be used to support very application-specific purposes. Figure 1 introduces our proposed methodology for federating models. This figure also includes steps to catalog models motivated by real-world use cases, and when possible, to evaluate and reuse previously cataloged models or standards-based domain models vetted by community experts. Furthermore, our approach only models use case scenarios taken from the messages exchanged between actors, as defined in each use case. Using this approach provides a straightforward way to federate models for business needs.

We hypothesize that because many of the messages hold data structures and information in common, users will not need to reinvent the wheel for every new event message encountered. Based on this, we further postulate there is a tremendous opportunity to build a catalog of federated models that can be reused not only to save time, but also to help standardize the way cyber-physical information can be exchanged in real-time adaptive decision support systems.

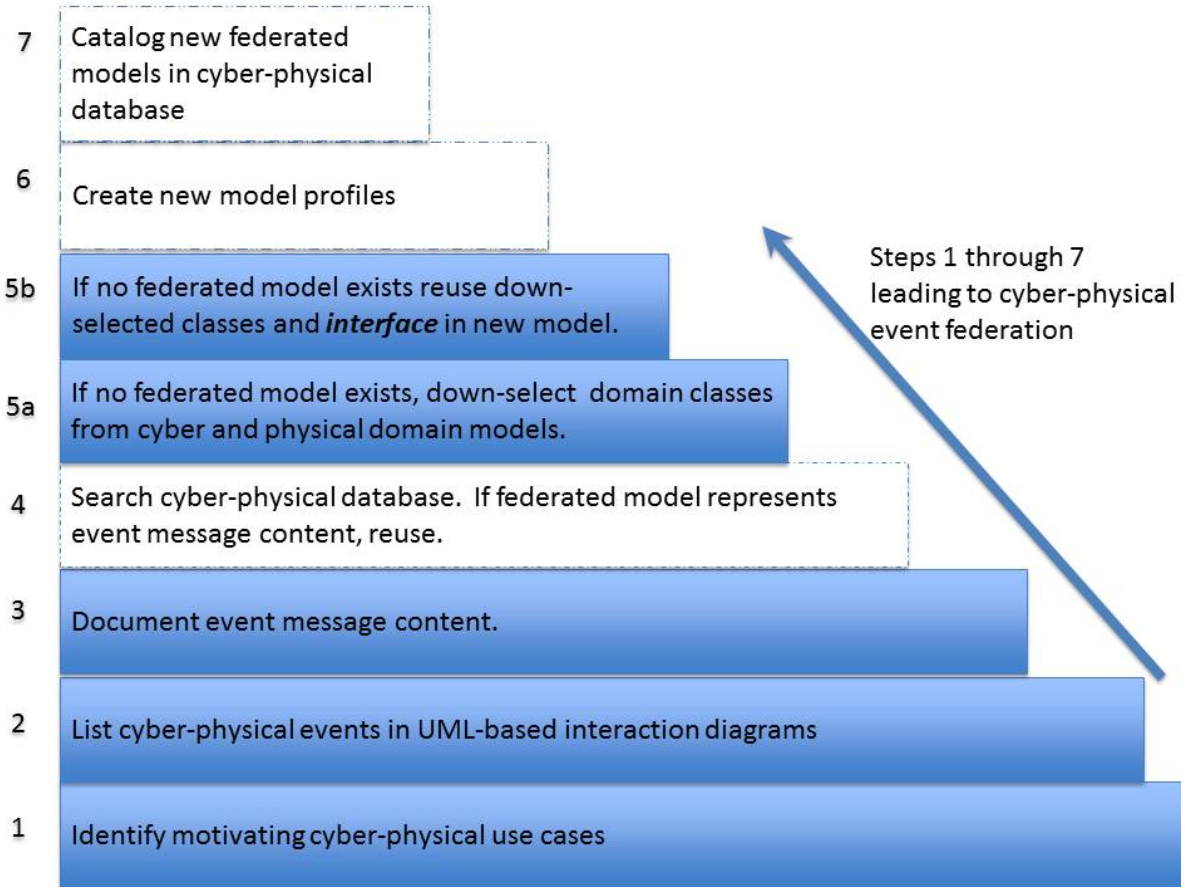

Figure 1. Proposed model federation methodology. Federated models are the product of identifying real-world use cases that provide details about event-driven information flowing between different actors. Given the nature of the use cases, expertise from both the physical (electrical grid) and cyber (computer infrastructure) communities must be employed. This report covers steps 1–3, 5a–b depicted in the diagram by shaded boxes. The dashed-line white boxes depict the next steps (4, 6–7) to be taken once federated models are identified and cataloged to support future searches for federated models.

1.2 Unified Modeling Language as a Modeling Standard

In a process control system environment, real-time scenarios require discrete facts and timely recognizable data in order to interpret values correctly and make quick decisions. Thus, the amount of time and number of processes needed to decipher and disseminate critical information must be limited. Machines, software developers, and systems integrators cannot properly interpret and use data without a syntactic structure or semantic context. Without a semantic or syntactic context, the string “7647-14-5” in a popular spreadsheet application could either be interpreted as the date “May 14, 7647” or a Chemical Abstracts Service number for sodium chloride. Observational data have similar potential ambiguities; for example, a temperature reading like “109.9” cannot be properly used and understood without such context as the unit of measure, geographic location (i.e., latitude, longitude) of the temperature gauge, and date and time the measurement was taken, as in, “101.9 Degrees Fahrenheit -119W 45N August, 19 2014

13:00." Over the years, data models have evolved as a general-purpose way of both semantically and syntactically describing machine-readable defined data structures. Once defined, these data structures are used in software applications and larger systems to temporarily cache state information, methodically pass messages between process interfaces (e.g., machine to machine), or store data offline. Ultimately, these data structures are used to (semi-)automatically accomplish a particular series of tasks.

As the approach of data modeling became standardized, the Unified Modeling Language (UML) provided a way to consistently specify the blueprints of a data model regardless of the programming language used to implement the data structures (see Figure 2). A key aspect of UML is that while it provides tools to model systems, it does not itself specify how to do so. Modeling is an iterative methodology, requiring expertise from knowledgeable modelers who are familiar with UML and existing information models, and from subject matter experts who can provide first hand guidance during the modeling process.

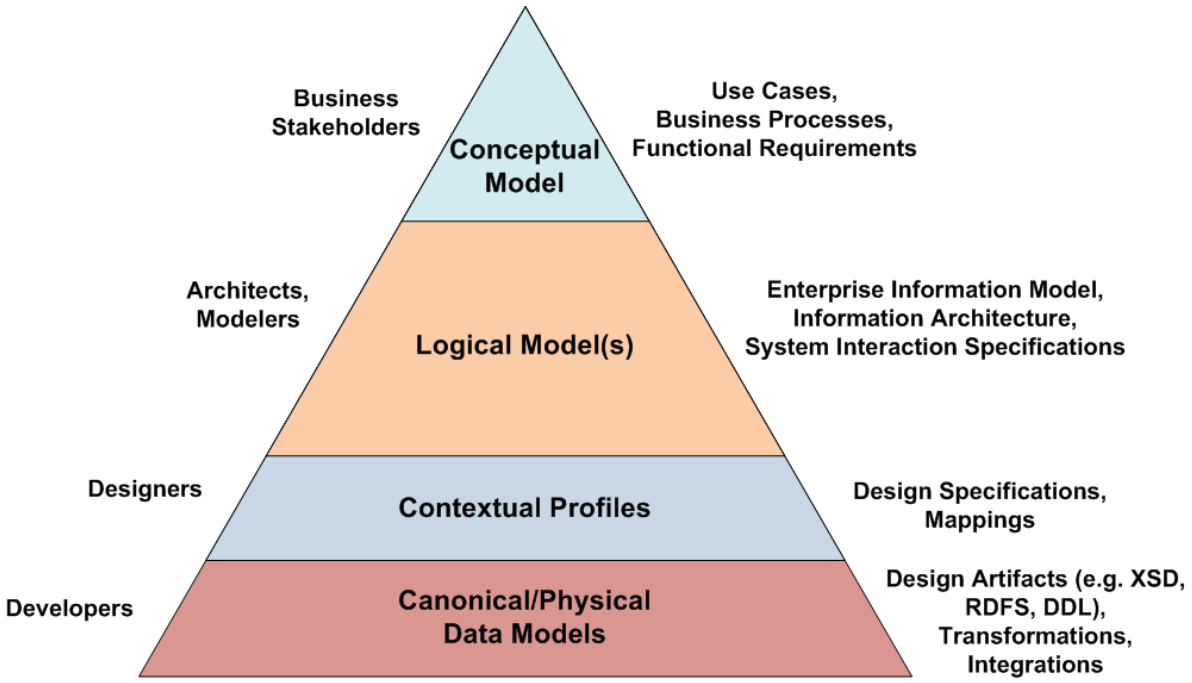

Figure 2. General modeling approach, where the process starts with conceptually matching initial requirements with use cases. Logical models are used to evolve the initial requirements into systematic representations that can be then be leveraged by designers and developers as data structures used in real-world software applications.

1.3 Document Roadmap

This report is organized to provide details on our approach to data federation and how data interfaces between models will ultimately be used to build UML profiles that can be more easily referenced and cataloged. Section 2.0 provides an extensively documented use case with well-defined messages that can show model federation logical perspective. This use case focuses on contingency analysis for an energy management system. Section 3.0 provides a more visionary use case with some interfaces requiring model federation. This use case features STEVE, or Security Threat EValuation Environment, our proposed suite of cyber-physical security evaluation applications that has the capability to identify cyber-attack scenarios in the power system Supervisory Control and Data Acquisition (SCADA) environment. While

the STEVE use case interaction diagrams have not been fully described, this case depicts how conceptual and logical models are defined in a similar way to the use case in Section 2.0, with additional details about how these models can ultimately generate UML profiles using CIMTool[1]. Section 4.0 discusses lessons learned and identifies next steps.

[1] CIMTool.org

2.0 Cyber-Physical Contingency Analysis

The following use case provides a description, an interaction diagram, an overview of the domain models used to support federation, and the class diagrams depicting federated logical model interfaces. The logical model interfaces were constructed by creating associations between domain models so as to reflect the overarching cyber-physical models.

This use case was studied by PNNL under the DOE Cybersecurity for Energy Delivery Systems project "Cybersecurity for EMS Decision Support Tools." More information about the algorithms and data needed for cyber-physical contingency analysis can be found in "Development of Cyber Aware Energy Management System Applications" (Sun et al. 2015) and "Cybersecurity for EMS Decision Support Tools Project: Technical Report" (Rice et al. 2015).

2.1 Use Case Description

Contingency Analysis (CA) is the portion of an Energy Management System (EMS) that system operators use to check for possible impacts to grid operation from an *N-1* condition; that is, the loss of a single component (generator or transmission line). The events considered within the scope of CA range from natural causes such as lightning strikes to malicious attacks. However, through cyber-attacks, the forced loss of more than a single system component becomes a possibility. It is not practical to compute the impacts from all *N-k* contingencies, as the number of combinations is huge. Our proposed solution for Cyber-Physical Contingency Analysis (CPCA) is targeted at mitigating this limitation.

The proposed CPCA application uses the following information as input:

1. *Substation vulnerability* – This information is obtained through periodic offline vulnerability analysis of the substation cyber (SCADA) infrastructure. This helps understand the "strength" of installed cyber defense mechanisms, and thereby provides an estimate of the most vulnerable substation networks in the system.
2. *Real-time cyber health* – Information about active targets in the SCADA network is obtained from network security devices that monitor the SCADA network infrastructure.
3. *Power system state* – This input is identical to traditional contingency analysis. That is, the current state of the power system is required to perform the "what-if" studies that will reflect the impact of changes to the system state.

Substation vulnerability information conveys the probability of compromise of the SCADA infrastructure within that substation. The real-time cyber health reveals if the substation is an active target. By combining this information, the CPCA identifies the grid components that could potentially be targeted. Thus, the CPCA application is able to drastically reduce the number of *N-k* combinations to be studied.

2.2 Interaction Diagram

The interaction diagram in Figure 3 is a graphical representation of the workflows of stepwise activities and interactions between users (also known as actors) and systems described by the use case in Section 2.1. Each arrow depicts (1) an interaction, (2) the source initiating the interaction and the intended target

recipient of the interaction, and (3) the context of the interaction. We interpret this interaction context as *message content*, where some type of structured data is being passed from the source to the target of the interaction.

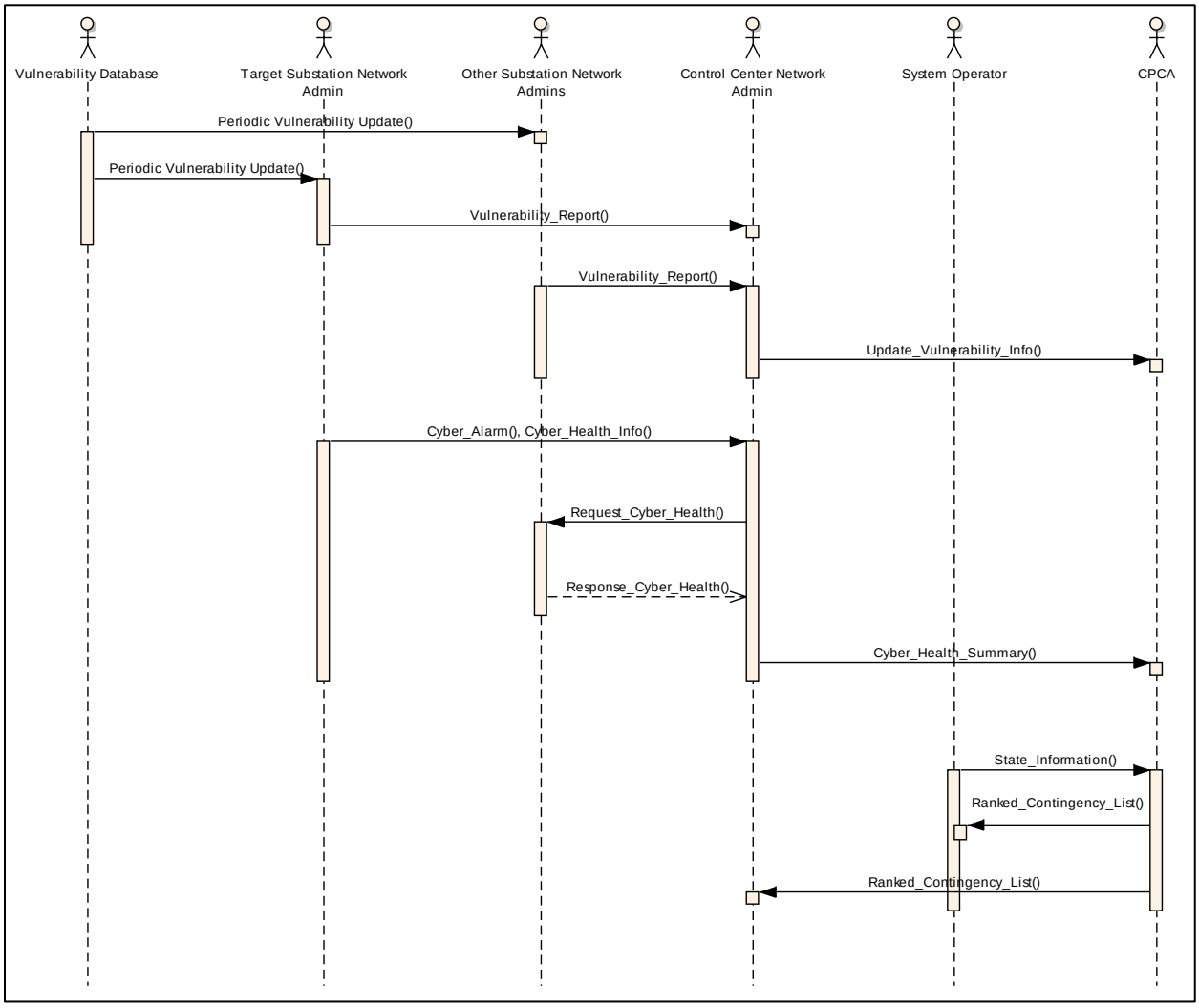

Figure 3. Interaction diagram depicting the CPCA use case

2.3 Event Message Content

Based on the use case interaction diagram, 12 messages were documented. Some of the messages use the same content, which provides an opportunity to reuse the same data model:

- *Messages 1 and 2 – Periodic_Vulnerability_Update()*
  Cyber network administrators across all substations perform annual vulnerability assessments of their substation. This includes the Target Substation Network Administrator (TSNA) and other Substation Network Administrators (SNAs). These network administrators use vulnerability and

threat information available in databases such as the National Vulnerability Database (NVD)[1] to compute the probability of a compromise in their corresponding substation.

- *Messages 3 and 4 – Vulnerability_Report()*
  The SNAs update their Control Center Network Administrator (CCNA) counterparts with the newly computed compromise probabilities.
- *Message 5 – Update_Vulnerability_Info()*
  The CCNA updates the CPCA application with the compromise probabilities. Weeks after the annual vulnerability assessment is complete, the TSNA notices alarms in the real-time cyber health information. He realizes that a potential compromise could impact grid operation and notifies the system operator and CCNA.
- *Message 6 – Cyber_Alarm(), Cyber_Health_Info()*
  The TSNA provides the CCNA with information on the cyber alarm that was triggered in the substation, along with the cyber health information of the substation. The cyber health information is envisioned to be a summary of the performance of the substation network during real-time operation.
- *Message 7 and 8 – Request_Cyber_Health(), Response_Cyber_Health()*
  Anticipating a coordinated attack on multiple substations, the CCNA requests cyber health data from other stations with a significant probability of compromise.
- *Message 9 – Cyber_Health_Summary()*
  The CCNA then passes this cyber health summary to the CPCA application.
- *Message 10 – State_Information()*
  The CPCA application also receives state information (line flows, etc.) from the system operator.
- *Message 11 and 12 – Ranked_Contingency_List()*
  Based on the cyber health summary and current power system state information, the CPCA application returns a ranked list of impactful *N-k* scenarios to the system operator and CCNA. With this ranked list of contingencies, the system operator and CCNA would be able to identify suitable mitigation strategies.

2.4 Domain Model Selection and Interface Requirements

Our approach for model identification has been thoroughly investigated in previous technical reports (Rice et al. 2015a; Rice et al. 2014). This approach includes cyber and physical threat indicators from STIX™-TAXII (MITRE Corporation 2012) (see also Figure 4), vulnerabilities identified in the National Vulnerability Database, as well as XML schemas.

[1] https://nvd.nist.gov/

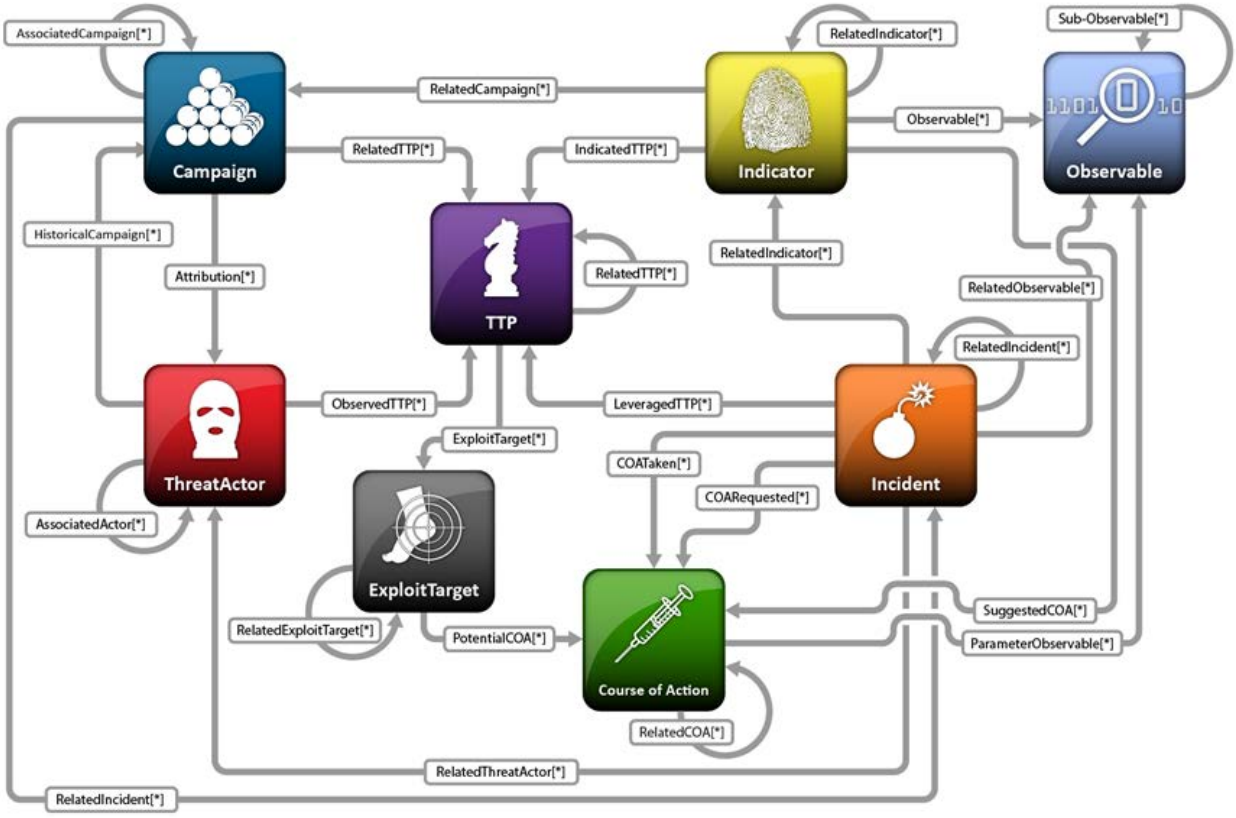

Figure 4. STIX™ architecture[1]

2.5 Cyber-Physical Contingency Analysis Federated Logical Model Interfaces

Evolving the messages defined in Section 2.1 into logical models requires research as to the availability of standard information models for each domain integrated. For this use case, the International Electrotechnical Commission Technical Committee 57 (IEC TC57)[2] and Distributed Management Task Force (DMTF)[3] Common Information Models (CIMs) were used. Our earlier studies found these models provided a rationale supporting most model component interface data structures that reflect both the cyber and physical domains. The data architect will use a modeling tool to create the necessary enterprise model. The enterprise model will reflect the information from the CIMs and any extensions, new classes, or new attributes to existing classes.

The following figures (Figure 5 through Figure 8) represent the messages in logical model form. Once in this form, it is possible to produce UML profiles which form syntactic and semantic specification of the message. Note that in several cases, the same interface model was used for multiple messages. The following figures are of the interface between models represented in UML; note that when multiple domain models are in the same figure, the different outline colors signify different models.

[1] https://stixproject.github.io/images/stix-architecture.png
[2] http://iectc57.ucaiug.org/
[3] http://www.dmtf.org/

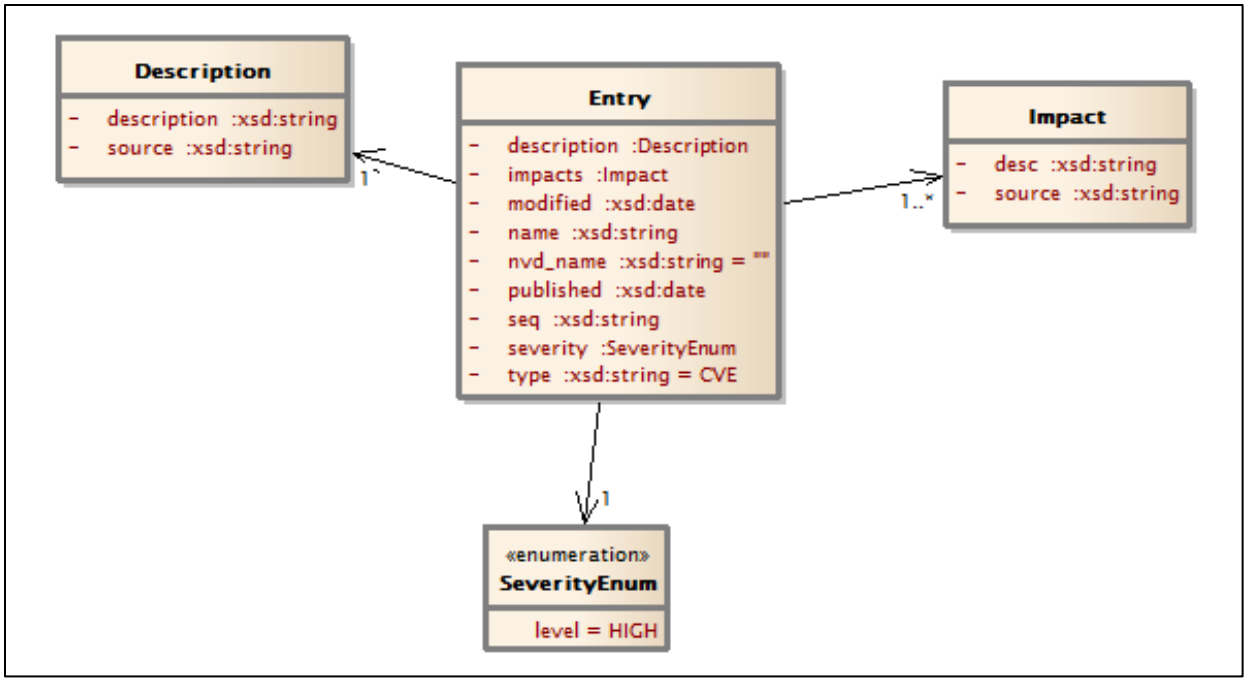

Figure 5. CPCA Use Case Messages 1 and 2. This UML represents the messages used in periodic updates of new records in the National Vulnerability Database.

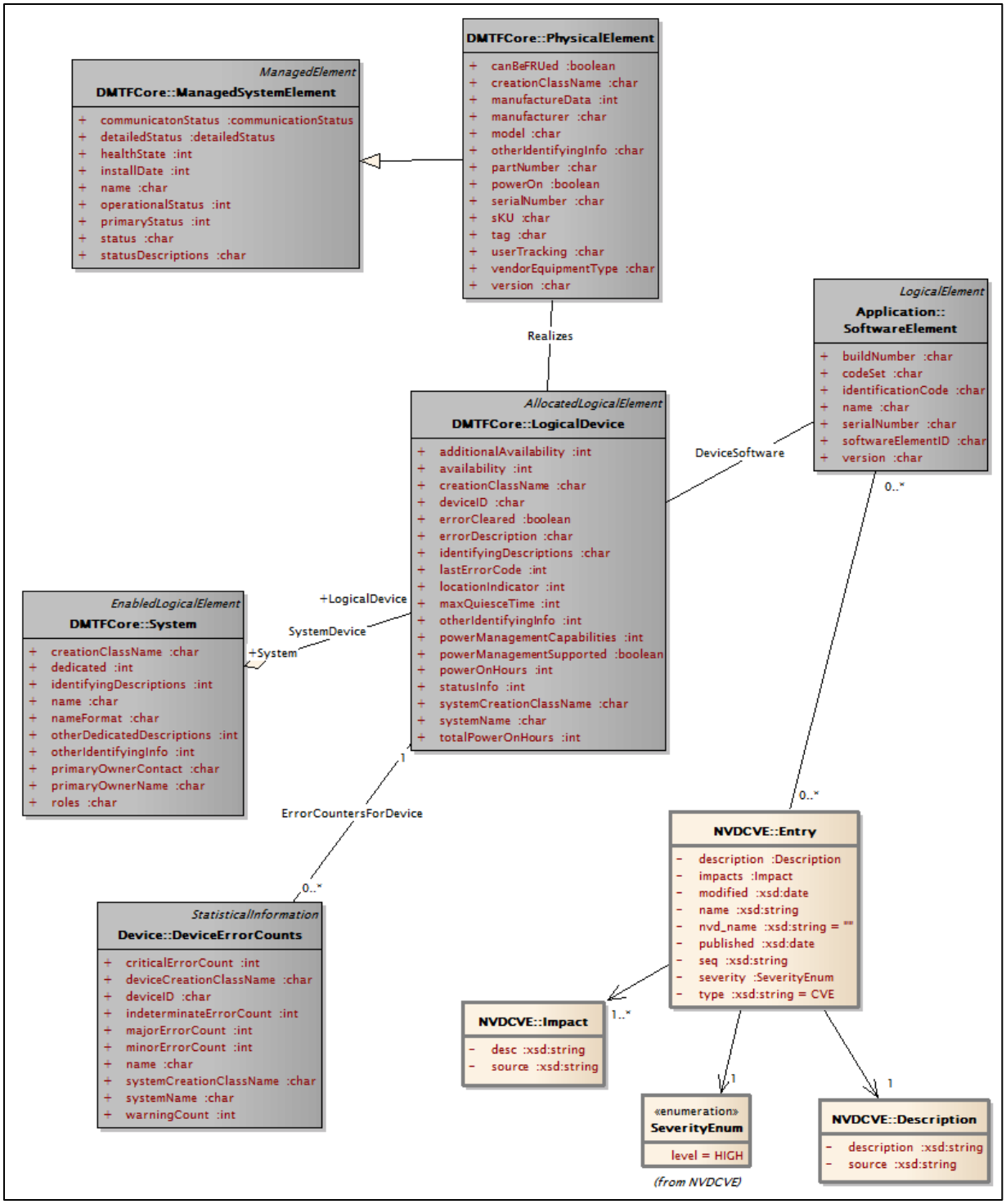

Figure 6. CPCA Use Case Messages 3, 4, 6, 7, 8, 11 and 12. This UML is of the interface between STIX™ (buff colored boxes) and the DMTF CIM (gray colored boxes). Note that messages 3 and 4 apply to vulnerabilities that impact specific equipment in the substation, whereas, Messages 6, 7, 8, 11, and 12 apply to cyber alarms, cyber health information, and ranked N-k scenarios (i.e., contingencies). This case shows how content can be used for multiple messages, thus reducing the complexity of the data model.

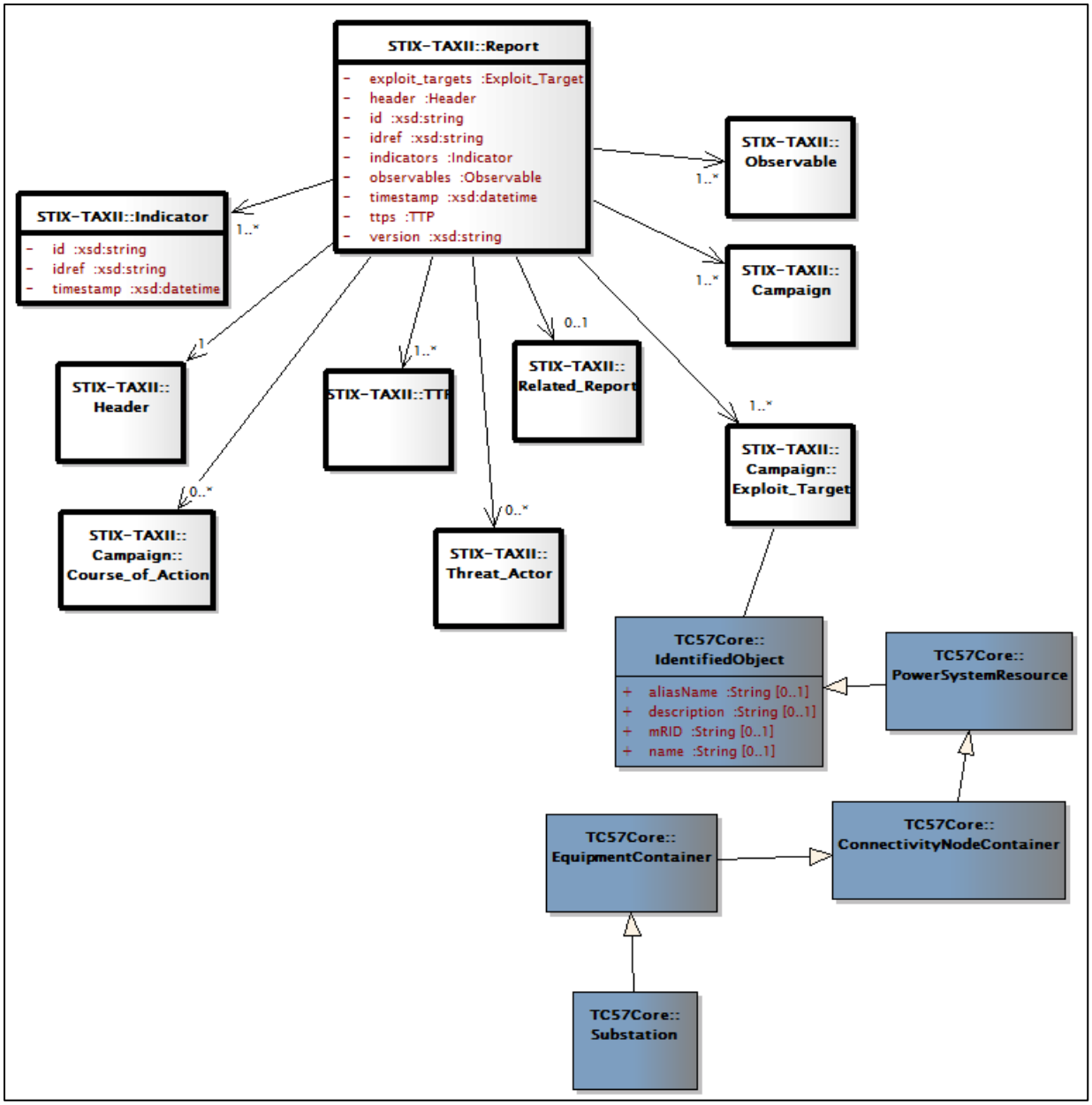

Figure 7. CPCA Use Case Message 5. The UML represents vulnerability information sent from CCNA using STIX™-TAXII (light gray colored boxes) and is aligned with TC57 CIM core components (blue colored boxes) modeled in CPCA.

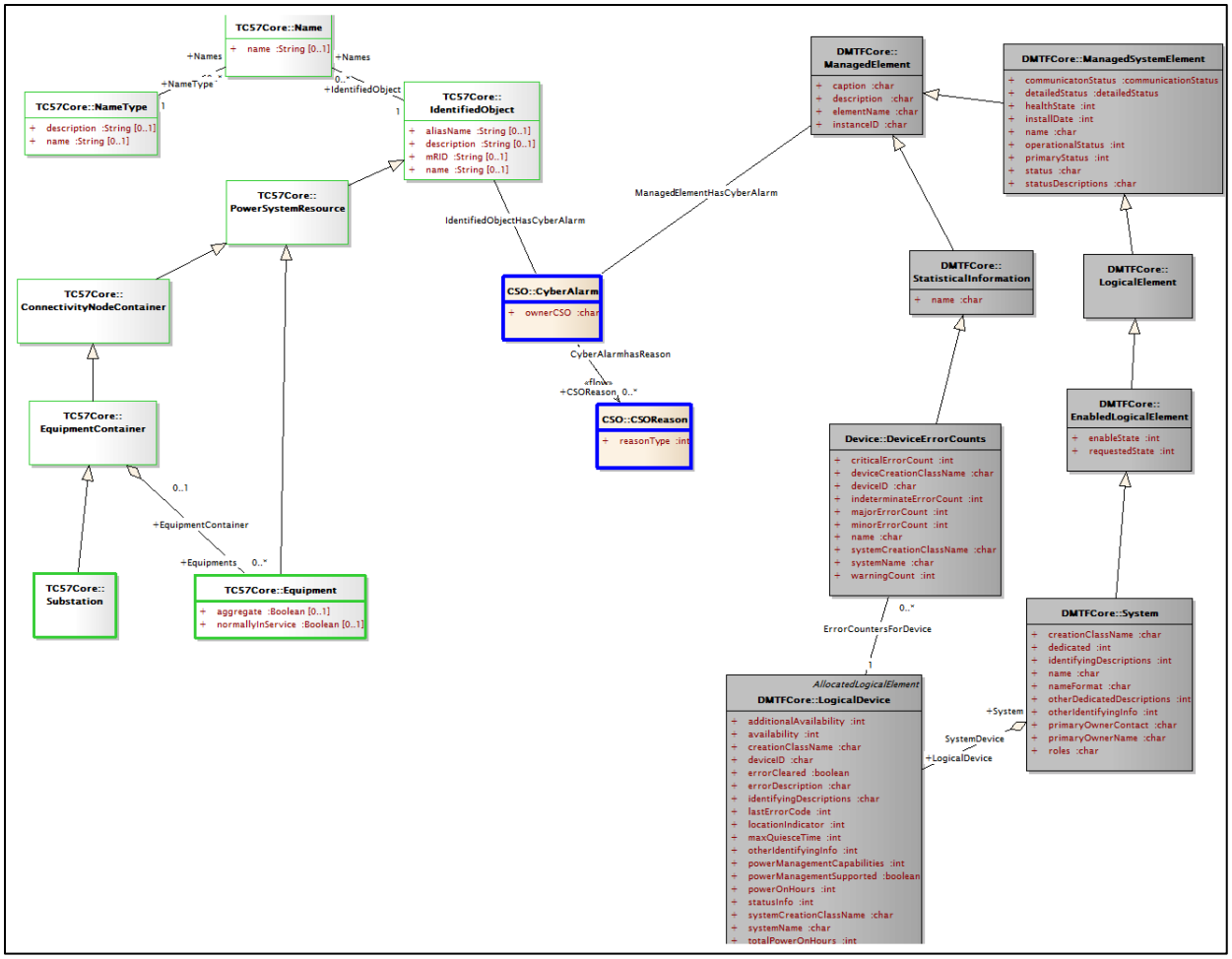

Figure 8. CPCA Use Case Messages 9 and 10. The UML represents the interface between CSO and CPCA in STIX™ and DMTF (dark gray colored boxes) for updating the CPCA with the cyber health summary for the substations, and between System Operator and CPCA in TC57 CIM (light gray colored boxes with green outline) for updating the CPCA with results from the state estimator.

3.0 Security Threat EValuation Environment Use Case

Similar to the contingency analysis use case in Section 2.0, the STEVE use case investigation provides a description, interaction diagram, overview of the domain models used to support federation, and the class diagrams depicting federated logical model interfaces made by creating associations between domain models to reflect overarching cyber-physical models.

As previously reported in "Model Federation for Enhanced Alarm Processing" (Rice et al. 2015a), STEVE is our proposed suite of cyber-physical security evaluation applications that has the capability to identify cyber-attack scenarios in the power system SCADA environment. The STEVE application is designed to correlate real-time cyber and power system data feeds to deduce possible connections between unexpected power system events and illegitimate cyber activity to identify cyber-attacks targeting power system operation. If a match is found, the system operators and network administrators are then able to trace the path of the attacker to potentially determine all compromised systems and/or power system data. An example scenario could be that a cyber-threat actor compromises a substation's cyber security and intentionally trips open a breaker. In this case, the SCADA system would generate an alarm indicating the change in state of the breaker. Similarly, cyber security technologies such as firewalls and intrusion detection systems would trigger corresponding alarms signaling malicious behavior in the cyber networks. The STEVE application would then correlate alarms from both domains to identify that the change in breaker status was a result of malicious cyber activity.

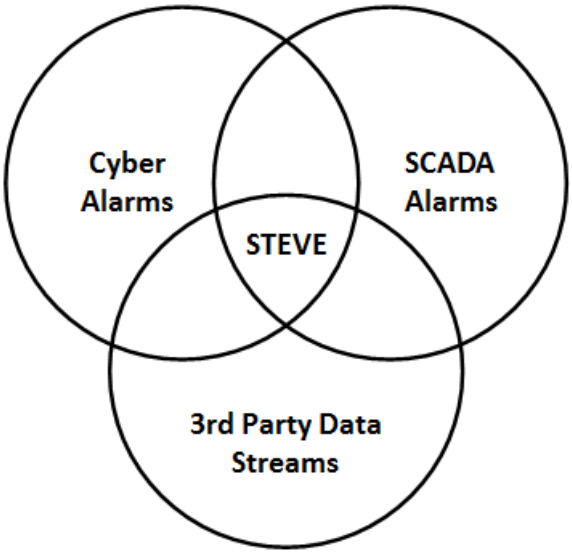

Figure 9. Security Threat EValuation Environment (STEVE)

The need for SCADA and cyber security alarms was illustrated with the example above, but data from third parties is also needed. Data streams from maintenance personnel will enable STEVE to identify abnormal situations arising from maintenance-type scenarios. Similarly, data streams from vulnerability databases will make STEVE aware of weaknesses in the cyber system, and thereby more cognizant of potential attack vectors. In this section, we explore a proposed STEVE use case in further detail. The interactions between different system actors in this use case highlight the need for model federation.

3.1 Use Case Description

After a conference where a PGO (Power Grid Operator) received a free USB storage device, he returns to his control center and plugs the device into a control center computer on a Friday. Rogue software on the USB drive installs malware on the human machine interface software on the PGO's PC. The PC is connected to the local area network behind the substation firewall. On this network, there are other computers to provide various SCADA functions. The malware begins looking for the OPC[1] server computer that it expects to see on the network and installs itself on the server through a demo user account. Over the weekend, the malware begins to "phone home" to the attacker who distributed the malicious USB device. The goal of the malware is to enable remote control of substation embedded devices through the OPC server. The unusual activities from the OPC server are recorded by network security tools. The following interaction diagram (Figure 10) describes the sequence of events following this malicious event.

Other cases in which STEVE could be used for improved coordination include the following:

1. The ability to quickly identify the assets impacted by a supply chain issue including discovered vulnerabilities and waterhole attack. The message from 3rd parties would come to the utility via a STIX™ message. The information within the message could then be correlated against power grid alarms (in IEC TC57 CIM) such as mis-operation of the circuit breakers, bad data, and generation not following AGC[2], to determine if an adversary is interfering with power grid operations.

2. The ability to inform operators of a Denial of Service (DoS) attack on a substation that is impacting SCADA communications (DMTF communication model). STEVE will align the SCADA alarms (IEC TC57 CIM) of stale data with SIEM[3] logs (STIX messages) stating the substation is under attack to help power grid operators understand.

3. The ability to mitigate a man in the middle attack in which an adversary is attempting to make the state estimation application solve to incorrect values. This done by identifying which measurements are taking an abnormally long time to be reported to the control center from a substation (DMTF communication model). With this information, the operator can choose to not use those measurements in state estimation (IEC TC57 CIM). The metrics used to determine the goodness of fit between the measurements and models (residual cost function) will indicate improvements in the state estimation without the impacted measurements.

Note these other use cases will not be discussed further in this report; however, they are representative of the improvements in grid operations that are possible with use of STEVE.

[1] OPC is the interoperability standard for the secure and reliable exchange of data. It is used in many industries that employ process control and automation systems. For more details, see https://opcfoundation.org/

[2] Automatic Generation Control (AGC) is a system that is used in electrical power systems to adapt to changes in load. AGC controls the output of several electrical generators at various separate geographic locations through commands sent from a central AGC command server.

[3] System Information and Event Management (SIEM). SIEM solutions track system logs, security events and alerts that are generated by hardware or applications on a computer network and analyze them in the context of security information (vulnerability reports, notices, etc.) from other sources. SIEM solutions may be implemented in hardware or software or sold as managed services.

3.2 Interaction Diagram

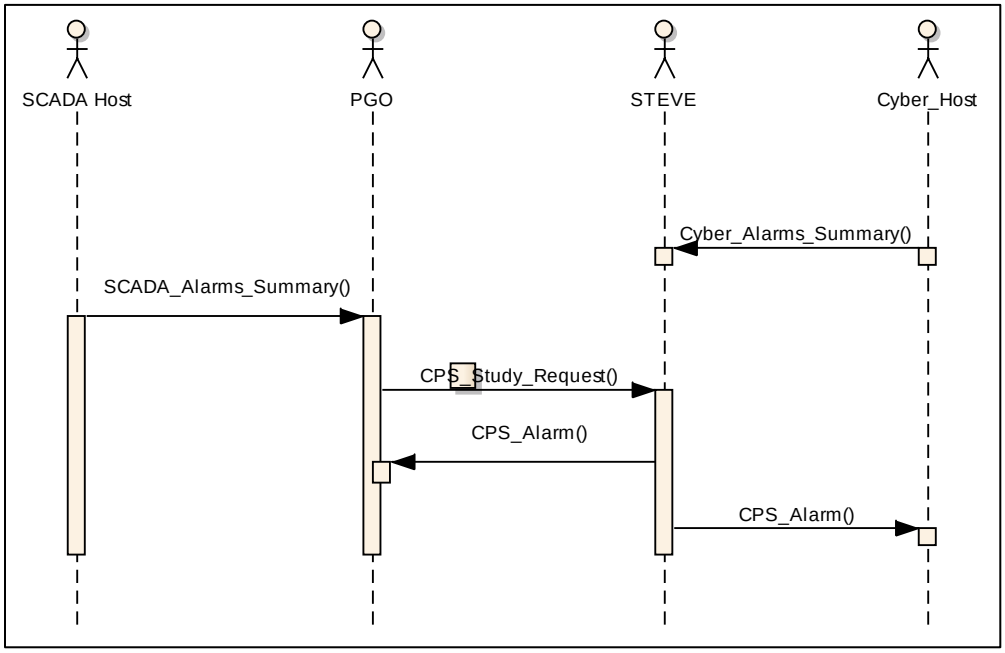

Figure 10. Interaction diagram depicting STEVE use case

3.3 Event Message Content

- *Message 1 – Cyber_Alarms_Summary()*
  The cyber system operator (CSO) reviews the unusual activities and begins investigating messages for other trends and related activities. These unusual activities are forwarded to STEVE.
- *Message 2 – SCADA_Alarms_Summary()*
  The PGO notices anomalous behavior (measurements) of substation equipment triggering alarms for the past hour.
- *Message 3 – CPS_Study_Request()*
  The PGO first runs a basic power flow analysis to cross-check three fifteen-minute snapshots of observed measurements. The PGO observes inconsistencies between the current system as revealed by SCADA and the expected system state. The PGO is able to narrow the problem down to an open line and is also able to confirm that the problem is not due to a physical grid event. In an attempt to narrow the root cause, the PGO engages STEVE and forwards the power flow analysis results.
- *Message 4 – CPS_Alarm()*
  STEVE analyzes the cyber security alerts provided earlier by the CSO against the anomalous behavior observed by the PGO. The events are identified as part of a cyber-attack. With the provided information, STEVE identifies all potentially compromised measurements and network devices. With this information, STEVE issues a cyber-physical system (CPS) alarm to both the CSO and the PGO.

3.4 Domain Model Selection and Interface Requirements

From the above example, it is clear that STEVE applications should have the capability to access information from multiple data models. This drives the need for data model federation. We envision STEVE having the capability to process the following types of information:

1. SCADA alarms (IEC TC57 CIM)
2. Cyber system alarms (DMTF and STIX™)
3. Third party data streams such as data from power system or cyber system maintenance personnel and vulnerability databases

3.5 STEVE Federated Logical Model Interfaces

For the purposes of this technical report, the logical model is still under development. The interfaces between DMTF, STIX™, and IEC TC57 CIM are shown in Figure 11 and Figure 12. The UML representation of Message 1 is shown in Figure 11. For Message 4 content in this use case (Figure 12), we are reusing the logical model from the contingency analysis use case (as shown in Figure 8).

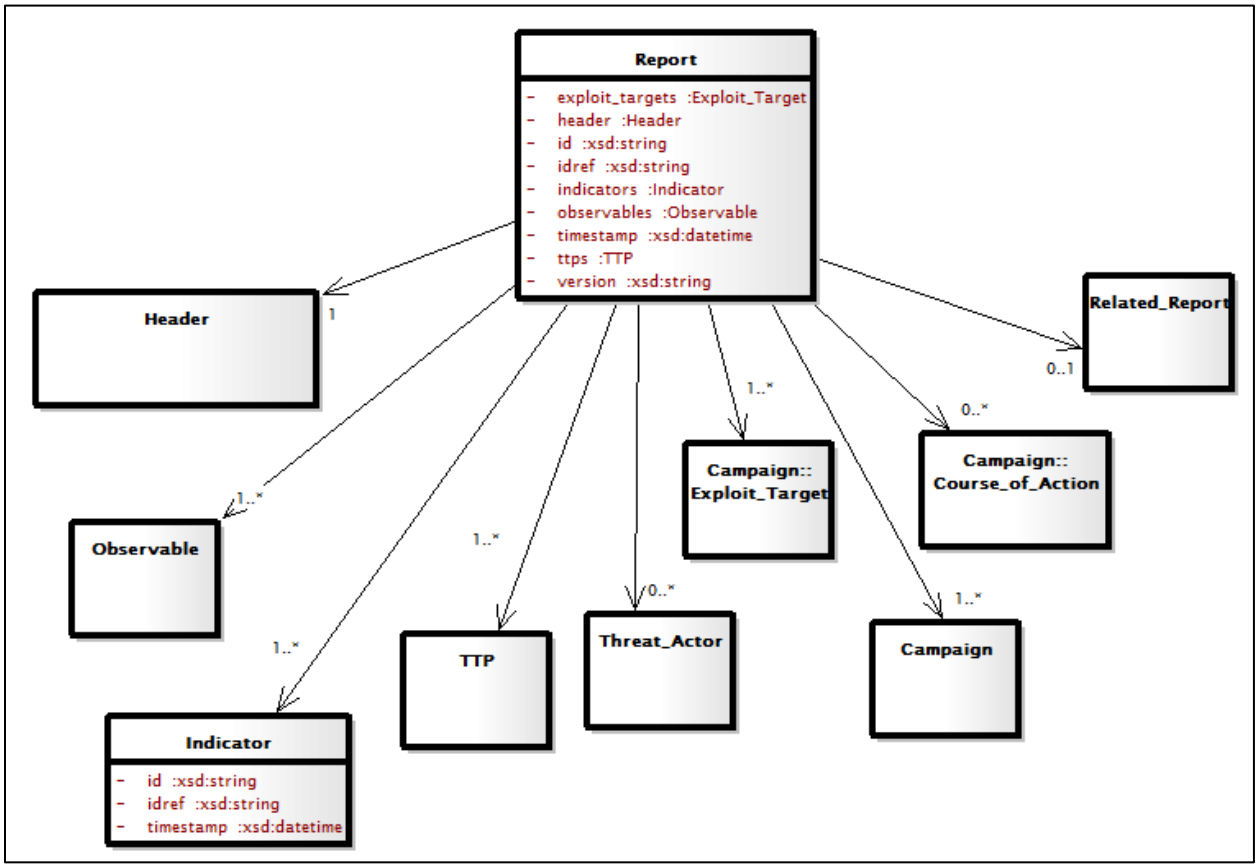

Figure 11. STEVE Use Case Message 1. This UML represents the messages used by the CSO to communicate information on unusual cyber activities. The classes are from the STIX™-TAXII model.

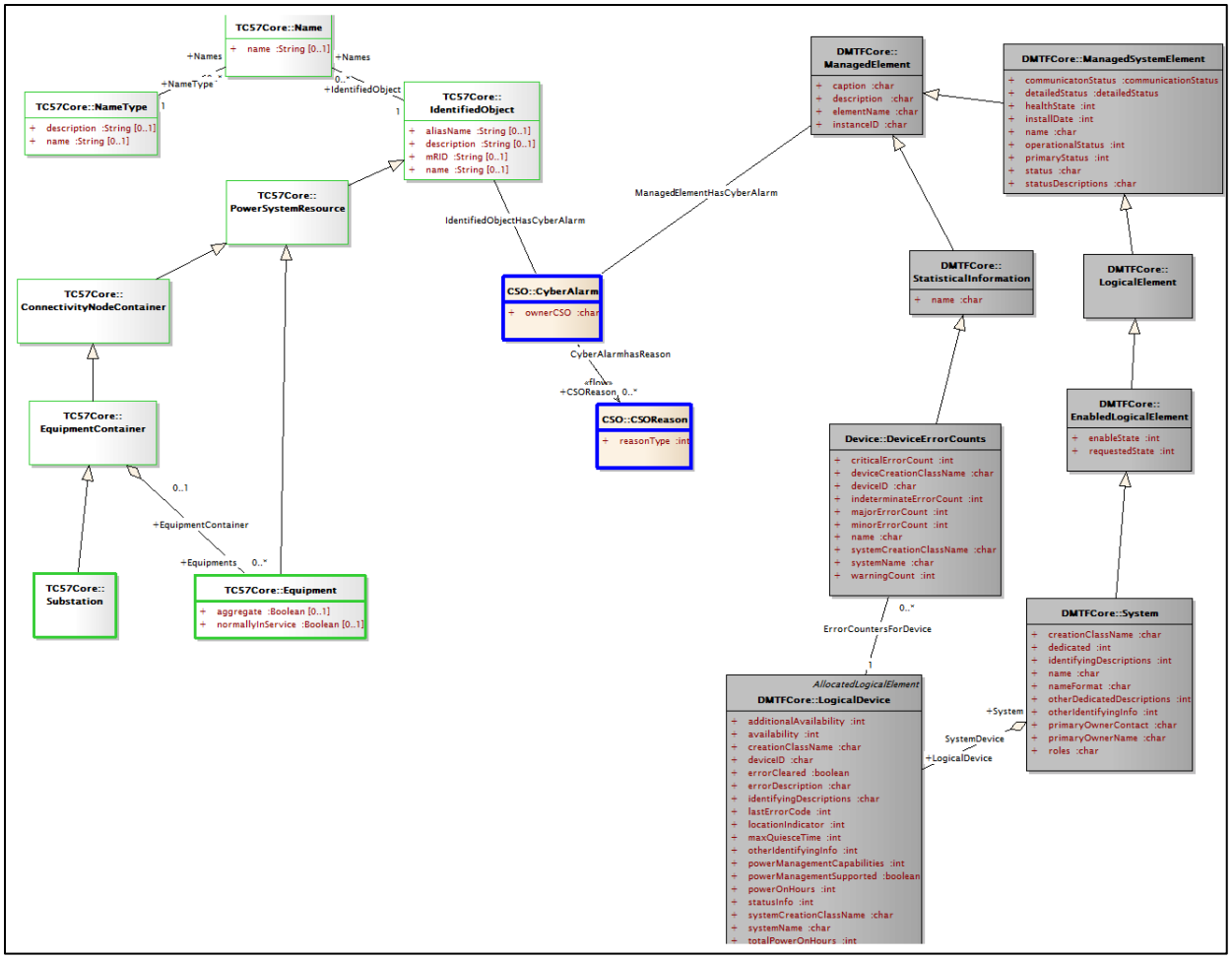

Figure 12. STEVE Use Case Message 4. This UML represents the messages used by STEVE to communicate the process measurements and system devices that it believes may be compromised by a cyber-attack. STEVE passes this information both to the CSO and the PGO. This model federation includes classes from STIX™-TAXII (buff colored boxes with blue outlines), DMTF (dark gray colored boxes), and the TC57 CIM (light gray colored with green outline).

4.0 Lessons Learned and Next Steps

4.1 Conclusion

Two different use cases were developed in order to effectively demonstrate the reusability of the interfaces. Such reusability is possible due to the fact that, although the profiles and message payloads will be unique to each business case or use case, the linking classes between modeling languages will be the same. For example, the link between the SCADA class in the IEC TC57 CIM and the computer class in the DMTF was the same in both use cases. More reusable interfaces will be discovered as more classes from separate models are linked together to address new use cases. The ultimate goal of this work is produce a catalog of the profiles that can be shared with others. This catalog will allow other developers to reuse concepts and repurpose links between models and thus reduce their development effort. In addition, the catalog will provide examples to other researchers who are building cyber-physical models; they can use an existing model in the catalog as the basis to create new, customized models.

4.2 Lessons Learned

Comparing the two modeling efforts, the following observations were made:

1. There is significant initial time investment required to identify domain-specific classes and federate models to support multiple domains.
2. Given the size and complexity of the domain models, current off-the-shelf technology used to examine the UML models is not sufficient. Top-down, stair step browsing for classes is highly inefficient. Search tools are needed to help orient users to domain model classes based on keyword searches and class definitions.
3. There are many situations where classes and federated models are reused. Based on these experiences, we can see that the publication of model federation use cases and development of federated logical models creates an overall long-term benefit in the cyber-physical community.

4.3 Next Steps

The next steps for this project will be focused on creating UML profiles. These canonical data models are used to diagram information exchange and provide common references. Using common information models, canonicals are defined using contextual profiles, where subsets of the enterprise model are leveraged. Each element of a profile can be traced back to a source element in the logical model (Enterprise Logical Information Model). A canonical data model is a physical data model, which is often realized as an XSD (XML Schema Definition) and becomes a Physical Design Artifact. Figure 13 depicts the overall modeling process that will be used to generate the profiles.

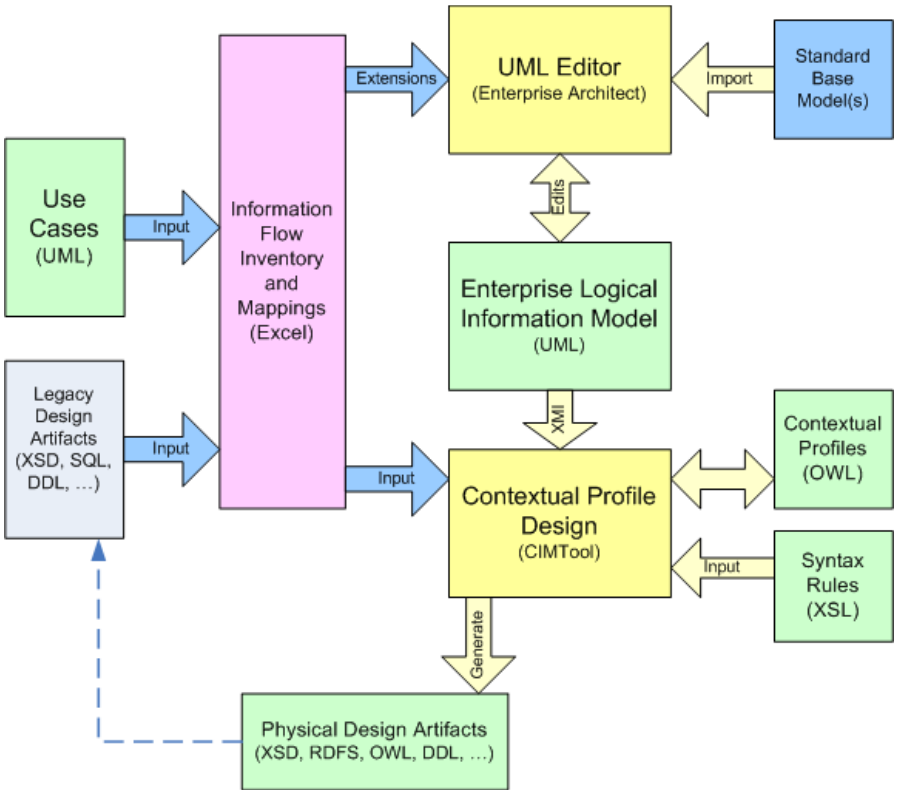

Figure 13. Overview of the modeling process

As mentioned above, the Contextual Profile Design describes the information exchange between the applications. Figure 13 depicts the use of the CIMTool to create and generate the XSDs for each information exchange. The first step in designing a Contextual Profile is to export the Enterprise Model as an XMI (XML Metadata Interchange) file, which can then be used in the CIM tool to select the information content (see Figure 14 through Figure 16).

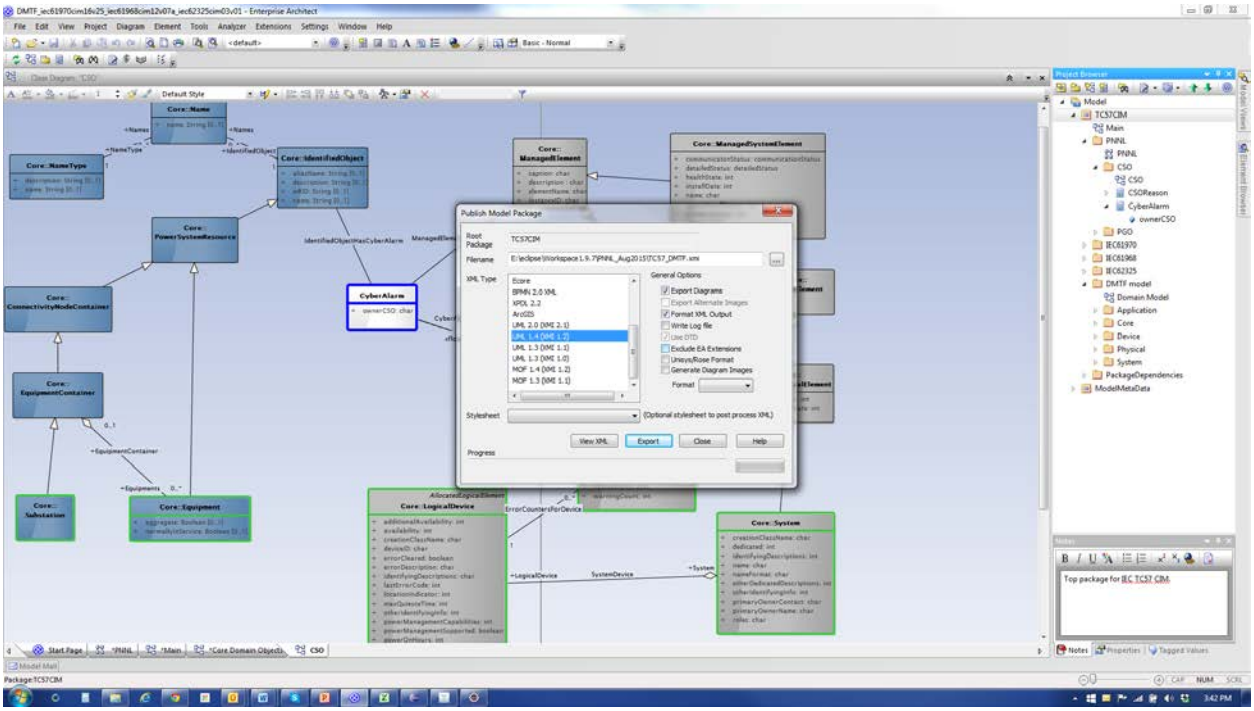

Figure 14. Exporting federated logical model into XMI form for the CIMTool to generate a profile

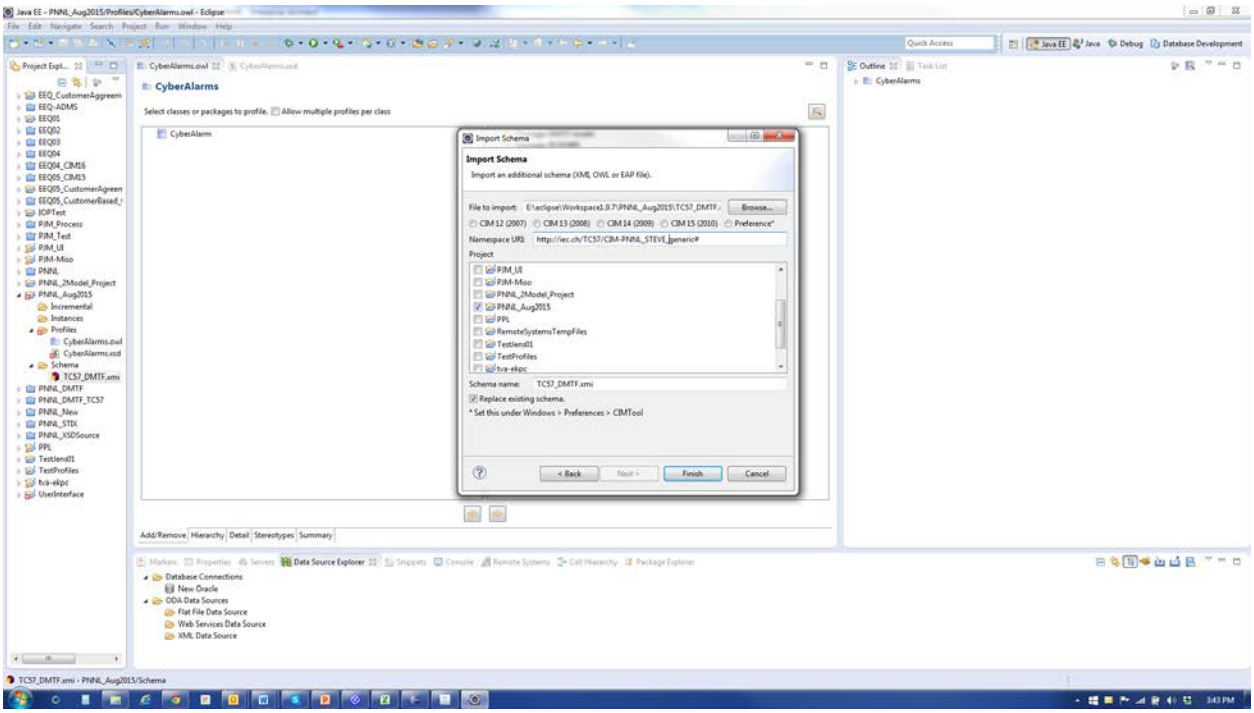

Figure 15. Importing the logical model into the CIMTool

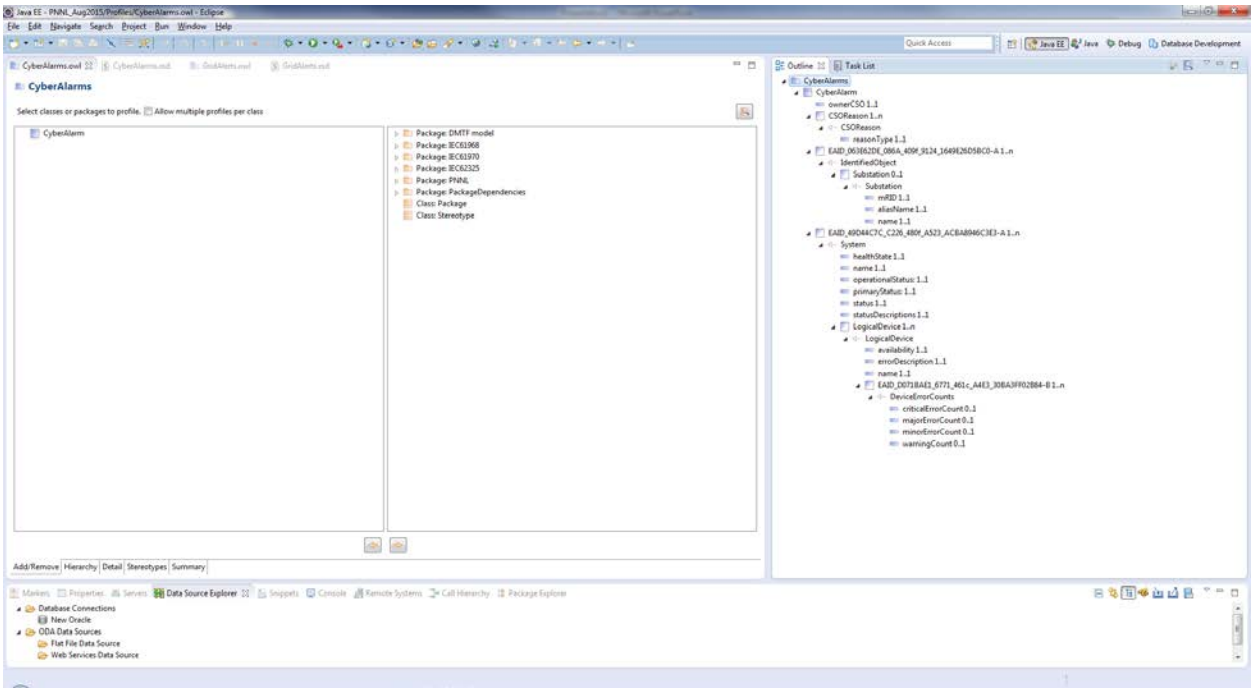

Figure 16. CIMTool will automatically store an XSD for the created profile

Once the UML profile is created, the profile can be cataloged in a database in a form against which queries can be performed.

Based on Figure 1, three steps need to be completed in the final phase of this project. Two of these are:

- Build profiles for model interfaces (Step 7 of Figure 1).
- Index the profiles and compile them into a semantic database that may be queried to find existing profiles (Step 6 of Figure 1).

Once the semantic database in Step 6 of Figure 1 is constructed, then future model federation efforts may query that database to search for existing model interface profiles that may be reused (Step 4 of Figure 1). Although developing a full search interface is outside the scope of this project, some query-by-example interfaces will be developed to show how it is possible to search for profiles in a more sophisticated way than is presently possible.

5.0 References

Barnum S. 2012. *Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX™).* MITRE Corporation.

Rice MJ, S Sridhar, EG Stephan, and RA Pawlowski. 2014. *Preliminary FEDSEC Use Cases.* Pacific Northwest National Laboratory, Richland, WA.

Rice MJ, S Sridhar, EJ Stephan, and RA Pawlowski. 2015. *Model Federation for Enhanced Alarm Processing.* Pacific Northwest National Laboratory, Richland, WA.

Rice MJ, S Sridhar, EG Stephan, Y Sun, and MR Vallem. 2015. "Cybersecurity for EMS Decision Support Tools Project: Technical Report." Pacific Northwest National Laboratory, Richland, WA.

Sun Y, S Sridhar, MJ Rice, and MR Vallem. 2015. "Developement of Cyber-Aware Energy Management System Applications." *2015 Grid of the Future Symposium*, Chicago, IL.

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

U.S. DEPARTMENT OF
ENERGY

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

www.pnnl.gov