CERT'S PODCASTS: SECURITY FOR BUSINESS LEADERS: SHOW NOTES

Using a Malware Ontology to Make Progress Toward a Science of Cybersecurity

Key Message: A common language is essential to develop a shared understanding to better analyze malicious code.

Executive Summary

"In 2011, the U.S. Department of Defense asked the JASON program to 'examine the theory and practice of cybersecurity, and evaluate whether there are underlying fundamental principles that would make it possible to adopt a more scientific approach' " [1, 2].

"The first JASON program report concluded that 'The most important attributes would be the construction of a common language and a set of basic concepts about which the security community can develop a shared understanding… a common language and agreed-upon experimental protocols will facilitate the testing of hypotheses and validation of concepts.' " [1, 2]

In this podcast, David Mundie, a member of CERT's Enterprise Threat and Vulnerability Analysis team, discusses the need for controlled vocabularies, taxonomies, and ontologies to make progress toward a science of cybersecurity. David describes his team's malicious code ontology as one example.

PART 1: WHY ONTOLOGIES ARE CRITICAL

Science of Cybersecurity

A report published by MITRE in 2010, titled "The Science of Cyber-Security" addressed the question: What would be needed to make cybersecurity into a science?

The first conclusion of this report was that the most important development would be the creation of "a common language and a set of basic concepts about which the security community can develop a shared understanding." [2]

In other words, what is needed is an ontology of cybersecurity. Once such an ontology exists, all other aspects of science, such as statistics and hypothesis testing, can build on this.

The Importance of Ontologies

An ontology is a hierarchical collection of standardized terms (controlled vocabulary), including the relationships among those terms.

The U.S. Defense Advanced Research Projects Agency (DARPA) and the World Wide Web Consortium (W3C) were early promoters of the importance of ontologies. The W3C produced a web ontology language called OWL.

Examples of ontologies include:

- International Classification of Diseases, 11th edition: the standard for describing medical conditions. It includes 68,000 diseases and their interrelationships.
- Google semantic network, containing 570 million objects and 18 billion facts.

PART 2: AN ONTOLOGY FOR MALWARE ANALYSIS AND COMPETENCY FRAMEWORKS

Background

The malware ontology has fewer than 300 terms so it not nearly at the scale of the examples cited above. The goal is to be able to reason about a formal representation of a knowledge domain, in this case, the analysis of malicious code.

This vocabulary is published in the report titled *The MAL: A Malware Analysis Lexicon*. Since the publication of this report, David's team has developed an OWL ontology using Protégé, which is Stanford University's ontology creation tool.

An Ontology-Based Competency Framework

Ontologies are now being used as the foundation for managing the competencies of a workforce. Ontologies support:

- understanding the domain and interrelationships within the domain
- reasoning about job descriptions
- reasoning in support of task analysis
- determining training needs
- determining the knowledge, skills, and abilities required of people tasked with finding, analyzing, and correcting malware, and performing root cause analysis

Steps to Develop the Malware Analysis Ontology

The steps taken to develop the malware analysis ontology are as follows:

- gained access to ten years of email from the CERT Malware Analysis Team, which produced 90,000 terms.
- eliminated those terms that occur in Moby Dick, presuming that if a term occurs here, it is not relevant to malware analysis. 70,000 terms still remained after this filtering.
- eliminated all terms that occurred only four times or less, which reduced the list to 5,000 terms.
- performed a manual analysis with subject matter experts, which resulted in 40 terms.
- reviewed textbooks and other reputable sources on malware analysis, increasing the list to 270-275 terms.
- encoded the terms into a structured dictionary based on the Internet Engineering Task Force's standard for dictionary servers.
- considered other SEI/CERT vocabularies such as those appearing in the Capability Maturity Model Integration (CMMI), the CERT Resilience Management Model (CERT-RMM), and CERT's insider threat work.

PART 3: ADDITIONAL SECURITY ONTOLOGIES IN DEVELOPMENT

Additional Ontologies in Progress

Additional ontologies are being developed for the following bodies of knowledge:

- insider threat detection
- insider threat modeling
- Computer Security Incident Response Team (CSIRT) coordination center information sharing
- incident response

Resources

[1] Mundie, David & McIntire, David. *The MAL: A Malware Analysis Lexicon* (CMU/SEI-2013-TN-010), Software Engineering Institute, Carnegie Mellon University, February 2013.

[2] MITRE. *Science of Cyber-Security* (JSR-10-102). MITRE Corporation, 2010.

Insider Threat blog: How Ontologies Can Help Build a Science of Cybersecurity, March 2012.

Mundie, David & Ruefle, Robin. "Building an Incident Management Body of Knowledge." Software Engineering

Institute, Carnegie Mellon University, 2012. Presented at the The First International Workshop on Security Ontologies and Taxonomies.

OWL Web Ontology Language, World Wide Web Consortium.

Protégé Ontology Development Tool, Stanford University.