Using SIRA Technology to Identify Cyber Vulnerabilities and Remediation

Chuck Rehberg, Chief Scientist

Semantic Insights™

A division of Trigent Software

October 3, 2011

Background

- In 2009, President Barack Obama declared America's digital infrastructure to be a "strategic national asset.“
- In May 2010 the Pentagon set up its new U.S. Cyber Command (USCYBERCOM), headed by General Keith B. Alexander, director of the National Security Agency (NSA), to defend American military networks and attack other countries' systems.
- USCYBERCOM is only set up to protect the military, whereas the government and corporate infrastructures are primarily the responsibility respectively of the Department of Homeland Security and private companies.

Problem

- In February 2010, top American lawmakers warned that the "threat of a crippling attack on telecommunications and computer networks was sharply on the rise."
- According to The Lipman Report, numerous key sectors of the U.S. economy along with that of other nations, are currently at risk, including cyber threats to public and private facilities, banking and finance, transportation, manufacturing, medical, education and government, all of which are now dependent on computers for daily operations.
- In 2009, President Obama stated that "cyber intruders have probed our electrical grids."

Pilot Program

- Pilot a collaborative environment, by combining technologies and services from our partners and ourselves, to support cyber planning and training.
- Employ the SIRA technology to help research cyber response scenarios from the internal DoD community (documented on their blogs/wikis), and/or the open cyber community.
- Make the results useful for planning and training.

The Research Goal

- The research goal is to use Natural Language Processing and semantic technology to find information relevant to:
  1. Specific cyber vulnerabilities
  2. How to close the vulnerability
  3. How to remediate an attack
- We propose to use the Semantic Insights Research Assistant (SIRA); a natural language and ontology-enabled research environment.

The SIRA Research Results

- SIRA's research results can be used by:
  1. Instructors to develop training materials
  2. Students to as part of exercises
  3. Planners to examine potential vulnerabilities of their infrastructure and plan response scenarios

Clues to cyber vulnerabilities in natural language text

In natural language text:

1. Sometimes vulnerabilities are identified as such.
   - *"The vulnerability could allow remote code execution if a user received a specially crafted WINS replication packet on an affected system running the WINS service."* - Microsoft Security Bulletin MS11-035 - Critical : Vulnerability in WINS Cou.... Retrieved on 01/09/2011 22:20:12, from http://technet.microsoft.com/en-us/security/bulletin/ms11-035
2. Sometimes an explicit situation giving rise to vulnerability is stated and not specifically identified as a vulnerability.
   - *"Certain domain names can allow execution of arbitrary code"* - Advisory: Certain domain names can allow execution of arbitrary code - Oper.... Retrieved on 01/09/2011 22:21:01, from http://www.opera.com/support/kb/view/938/
3. Sometimes only the symptoms of a situation giving rise to a potential vulnerability are described.
   - "*this would enable remote command execution on machines running compromised versions*." - Ivan Fratric's Security Blog: WordPress source code compromised to enable r.... Retrieved on 02/09/2011 22:05:33, from http://ifsec.blogspot.com/2007/03/wordpress-code-compromised-to-enable.html

Clues to remediation of cyber vulnerabilities in natural language text

In natural language text:

1. Sometimes remediation is called out as such.
   - "*Ensure that the allow URL_fopen is disabled on the web server to help limit PHP vulnerabilities from remote file inclusion attacks.*" - US-CERT Technical Cyber Security Alert TA11-200A -- Security Recommendation.... Retrieved on 02/09/2011 14:13:03, from http://www.us-cert.gov/cas/techalerts/TA11-200A.html
2. Sometimes an explicit remediation to vulnerability is stated and not specifically identified as a vulnerability
   - "In BIOS you must enable hardware virtualization support (VT) Trusted eXecution Technology (TXT) and VT-d." - Flicker: Minimal TCB Code Execution. Retrieved on 02/09/2011 22:09:24, from http://sparrow.ece.cmu.edu/group/flicker.html

All the previous examples were generated by SIRA

- To do this I posed the following investigation to SIRA:
  - "What allows remote code execution?"
- An existing dictionary and ontology was identified for IT terms, concepts and relationships.
- SIRA was directed to research:
  - CERT vulnerability notes database at http://www.kb.cert.org/vuls/ dating back to 2000 with about 2700 entries.
  - A technical alerts catalog at http://www.us-cert.gov/cas/techalerts/ that appears to go back to 2004.
  - The Internet (using Google to identify a large set of the most likely documents for SIRA to read)
- SIRA generated natural language reports with Bibliography

Accuracy and Completeness of the Results

- Using tools provided by SIRA, Researchers/Domain Experts can increase the accuracy and completeness of the results by:
  1. Describing to SIRA a set of potential threat scenarios
  2. Adding any missing domain-specific terms (senses and synonyms) to the dictionary
  3. Providing “implication rules” to identify references that can infer a vulnerability.

SEMANTIC INSIGHTS™

A Division of Trigent Software, Inc.