

2013 DHS S&T/DoD ASD (R&E) CYBER SECURITY SBIR WORKSHOP

Linguistic Analysis of Malware

Charles River Analytics

Dr. Terry Patten

Catherine Call

July 24, 2013

This material is based upon work supported by The United States Air Force under Contract No. FA8650-11-C-1052.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of The United States Air Force.

APPROVED FOR PUBLIC RELEASE: 88ABW-2013-3151



**Homeland
Security**

Science and Technology



Company Profile

- Small business headquartered in Cambridge, MA since 1983
- 100+ employees and associates
- Long-term collaborative partner with world-class universities
- Recognized science & technology (S&T) developers across DoD
- Strong participation in professional societies/panels
- TS-cleared personnel and facility
- GSA schedule for IT services
- Service Divisions
 - Decision Management Systems
 - Cognitive Systems
 - Sensor Processing & Networking

charles river analytics

Customer Need

- Cyber attacks pose a significant threat to US security
- We need to improve the ability of software systems to detect sophisticated cyber attacks, and then act quickly enough to either prevent or remediate the attack
- Our Phase II Software Protection through Autonomic Knowledge Representation (SPARK) effort focused on detecting
 - Nation-state class attacks
 - Novel attacks

Approach

- The science of linguistics has developed successful techniques for capturing generalizations in sequential data
 - People easily understand sentences they have never heard before
 - Need systems that understand attacks they have never seen before
 - Key in each case: a grammar
- Grammars capture generalizations
 - Allow us to match novel attacks
 - Less rigid than a signature-based approach
- Functional (rather than structural) grammar
 - Encodes the functions of malware and the structures that realize them
 - Form follows function!
- Learning
 - The system can learn (with human guidance) from new attacks and quickly update the grammar to prevent similar attacks in the future



Technical Approach



- Input observed features from
 - dynamic analysis
 - static analysis
- Parse observed data using a functional attack grammar
- Output the functional characterization of the malware
 - The attack goals
 - How the attack works
- Tiered approach to software protection
 - Automated defense techniques to function autonomously
 - Fall-back to human involvement when anomalies are encountered
- Key technologies
 - Natural Language Processing (NLP)



Benefits



- Automatic, in-depth characterization of malware
- Automatic analysis of *novel* malware
- Helps human analysts find interesting behavior



Current Status

- Completed Phase II SBIR
- Working full-scope prototype
- Ready to explore real-world data and applications



Next Steps



- Evaluation on pilot applications
- Explore integration into cyber security environments/tools
- Seeking partnerships, licensing opportunities for
 - Government transition
 - Commercialization

Contact Information

Dr. Terry Patten
Principal Scientist
617.491.3474 Ext. 582
tpatten@cra.com

Catherine Call
Senior Software Engineer
617.491.3474 Ext. 575
ccall@cra.com

charles river analytics

- + Charles River Analytics Inc.
- 625 Mount Auburn St.
- Cambridge, MA 02138
- p: 617.491.3474
- f: 617.868.0780
- www.cra.com