CYBER SECURITY DIVISION

2013 PRINCIPAL INVESTIGATORS' Meeting

Cyber Security and Big Data: The Role of CCICADA

CCICADA Center, Rutgers University

Fred S. Roberts, Director

September 17, 2013

Homeland Security

Science and Technology

Team Profile

- CCICADA is the Command, Control, and Interoperability Center for Advanced Data Analysis
  - A DHS University Center of Excellence
  - Based at Rutgers University, with 16 academic and private sector partners, and with a sister center, VACCINE based at Purdue
- We build mathematical and computational foundations to extract both **knowledge** and **practical consequences** from massive, complex or unstructured data.
- Product: **Powerful analytical tools** for information sharing, collaboration, and data-driven decision support.
- We have extensive experience in **developing theory-based educational programs**

Customer Need

- Cyber attacks are massive:
  - They come from hundreds of thousands of attack IP addresses (300,000+)
  - They generate billions of total page views (7+ billion)
  - They drive data rates of 125 Gigabits every second
  - They involving thousands of servers (2500+)
- Cyber attacks are a high-speed threat:
  - Pervasive (1M+ hops/sec)
  - Persistent: Minutes to hours (>100Gbp/sec in 4 hrs.)
- The Need:
  - **Build on data analytics to create real-time applications**
  - **Cyber security education for every kind of user**

Approach: Building on Data Analytics to Create Real-time Applications I

Tools for Detection of Attacks

- **Machine learning** to detect Chinese censorware
- Detecting cross-site request forgeries
- Detecting interdomain routing anomalies
- Detecting web-based attacks in which attackers surreptitiously inject code into HTTP requests.
- Anomaly detection for large networks using **large graphs** that capture interactions between hosts and failed domain names.
- **Bio-inspired** distributed decision making for anomaly detection
- Using **natural language processing** to detect chatter about a cyber attack.

Approach: Building on Data Analytics to Create Real-time Applications II

Tools for Attack Prevention/Mitigation/Response

- **Cyber obfuscation***:* Increase confusion & uncertainty of attackers; redirect their resources to minor targets
  - By randomization or replication or modulation
- NetMelt to find k best edges to remove to minimize virus propagation in network
- Developing complex systems with **self-healing properties**
  - Application to self-healing SCADA systems (smart grid)
- Related: hardening communication security of energy delivery systems (work with Detroit Edison and EPRI)
- Cloud security: Preventing cloud administrators from snooping

Approach: Building on Data Analytics to Create Real-time Applications III

Privacy and Secure Information Sharing

- Privacy-preserving data analysis – using **cryptographic approaches**
- Secure and private database access: Neither server nor third party should learn what the client's queries are
- **Secure multiparty computation**
- **Data anonymization** to enable sharing
- Automated methods to determine trustworthiness of online sources
- Constrained trustworthiness models in disaster situations

Approach: Building on Data Analytics to Create Real-time Applications IV

Human in the Loop

- Tools to help Android users better understand application permissions
- Targeting naïve or careless users of devices: **Crowdsourcing** for threat detection
- Botnet detection: **Using biometrics** of key strokes to identify infected machines

Approach: Cyber Security Education for Every Kind of User

Cyber Security Education

- DHS has 200,000 employees; HSE has millions.
  - Governments need cyber security education programs
  - So does the private sector
- The public use of sophisticated tools creates vulnerabilities
  - The public needs cyber security education
  - Education should begin in high school, or earlier
- **Challenge: Education moves slowly; cyber threats and defenses move very rapidly**
- The Project:
  - Survey the field
  - Recommend programs of education for DHS, HSE, private sector, and the public
  - Theory-based approach: Recommend research needed to learn which educational approaches work best for which audiences

Benefits

- CCICADA builds on powerful data-analytic advances: machine learning, large graphs, natural language processing, cryptography, obfuscation
- Our prototype tools have been tested on real or simulated data and are ready to be transitioned for practical use
- Our tools recognize the problems created when humans use sophisticated devices
- Our proposed educational programs
  - will be developed rapidly
  - will build on programs already under development in academia, the private sector, and government

Current Status

- **Building on Data Analytics to Create Real-time Applications:**
  - Many of the projects mentioned above have developed tools either already in prototype form or close to it.
  - However, all will benefit from more research
  - All need our assistance to transition to practical use.
- **Cyber Security Education For Every Kind of User:**
  - Project just beginning
  - Mini-workshop with experts in early Fall; later, leads into large cyber security education workshop in Jan.
  - Survey initiated in Sept., completed by Dec.
  - Recommendations for educational programs and research on what works in cyber security education delivered by Dec.

Next Steps

- **Building on Data Analytics to Create Real-time Applications:**
  - Further exploration of natural language processing for attack detection, crowdsourcing for threat detection, & more
  - Understanding fundamental theoretical concepts that cut across tools and applications
- **Cyber Security Education For Every Kind of User:**
  - Incorporate proven concepts of mathematics education
  - Extensive discussions with those developing and testing new programs, leading to recommendations
  - Special emphasis on educating people to deal with **cyber threats on devices no one can conceive of today** and corresponding **cyber defense tools not yet invented**

Contact Information

Fred S. Roberts, Director of CCICADA
froberts@dimacs.rutgers.edu, 848-445-4303

Researchers Involved in Projects Discussed:

- CMU: Jaime Carbonell, Eugene Fink, Virgil Gligor, Ed Hovy
- Rutgers: Dennis Egan, Tina Eliasi-Rad, Nina Fefferman, Vinod Ganapathy, Paul Kantor, Muthu Muthukrishnan, Bill Pottenger, Rebecca Wright
- Stevens: Susanne Wetzel
- UIUC: Dan Roth
- Applied Communication Sciences: Ashish Jain
- AT&T Labs Research: Vladimir Kolsenikov, Abhinav Srivastava
- Bell Labs: Jin Cao