

# Bio-Inspired Anomaly Detection

## Cyber Security Division 2012 Principal Investigators' Meeting

10/11/12

**S. Raj Rajagopalan**  
**Scientist**  
**HP Labs/Honeywell**  
**Sraj.raj@gmail.com**  
**908-305-1681**

# Bio-Inspired Anomaly Detection

- Addressing TTA 13 – Nature-inspired Cyber Health:  
**“Bio-Inspired Distributed Decision Making  
for Anomaly Detection”**
- PIs:
  - Prof. Nina H. Fefferman, Rutgers University,  
Ecology, Evolution, and Natural Resources, and DIMACS/CCICADA
  - S. Raj Rajagopalan, HP/Honeywell
- Team: Two Post Doctoral Researchers, a couple of grad students, a research programmer, and some very smart people to ask for advice as we go.
- Collaboration between the NJ and AZ researchers mostly by electronic communications, with infrequent in-person visits

# Technical Approach

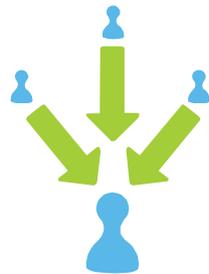
- Problem to be addressed: **Early/Accurate Detection of Threats on Widely Distributed Computer System Networks**

- Traditional methods utilized:

Centralized knowledge

&

Defined-pattern matching  
(or violation)



- More recent methods utilized:

Distributed knowledge

&

Known When Seen



(like pornography)

no picture available

# The Problem: What is new

- Detecting malicious activity (malware, insider threats, etc) is getting (much) harder
  - Systems increasingly complex
  - Adversary increasingly sophisticated and fast
  - Environments getting more distributed than ever before
- E.g. HP's IT network
  - Has ~ 200K computers
  - Has ~10K networking equipment
  - Spread over four continents
  - Managed from three global hubs (data centers)

# The Problem

## Specific challenges

- Data volumes
- Heterogeneity of Technology and Attacks
- Adaptable adversary

Threat detection is distributed and resists prepackaged solutions.

# The Approach

- Routers and switches today
  - Are distributed in the network
  - Have significant CPU, memory, and storage resources but no significant use for them

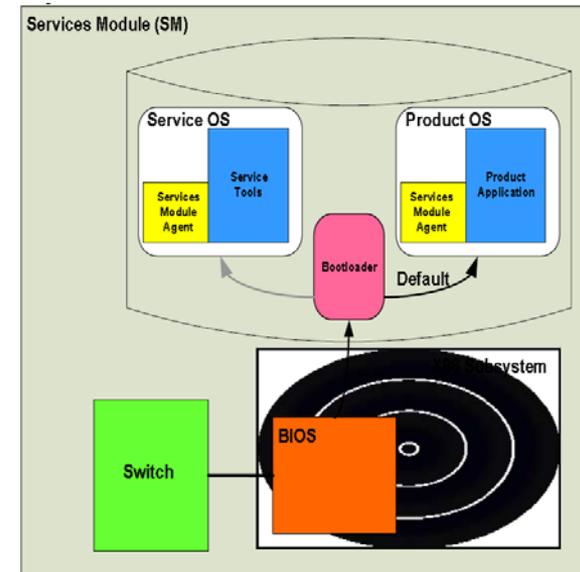
What purpose can we put these resources to?

- Distributed Anomaly Detection
  - Use in-network resources to store and analyze local traffic
  - Share significant events

# The Approach

## Leveraging infrastructure

- The ONE Blade on the HP Pro Curve Switch
  - Capable of running an independent computer with local CPU, memory and storage
  - Full access to traffic
  - Independent control plane
- Anomaly Detection
  - ONE blades can store and analyze local traffic
  - Communicate with HIDS on local networks
  - Run anomaly detection on local traffic
- But distributed anomaly detection is hard!



# Nature provides model systems

- Anomaly detection in nature is a well-studied problem and there are a few perfect examples that involve both Distributed knowledge & 'Known When Seen' threat identification.
- We will focus on two that utilize/favor detection of different types of anomalies on different system architectures:



- Both Bees and Ants rely for their survival on their ability (as colonies) to detect both known and unknown anomalies – Natural Selection has produced these populations that employ highly efficient distributed algorithms to identify predators, food sources, nest sites, etc.

# The Mathematics of Social Insect systems

- Each individual functions as an independent gatherer and analyzer of data, but can share conclusions with others, solicit the conclusions of others, and choose what (if any) data to share with which others. Conclusions are drawn using direct evidence, past experience, and the input of neighbors.
- Mathematical models capturing some of these dynamics already exist within the biological literature, but will require re-working to be able to apply them to computer system networks
- Evidence of performance of these natural systems in shifting real-world environmental conditions leads us to believe that these will be more effective at threat detection, without requiring prior models of threat types, and without requiring organizational centralization of information collection and analysis

# Moving Pieces: Exploring how these algorithms perform on different networks

Algorithm Type	Network Topology	Anomaly Type
Bee-Inspired Algorithm	Scale Free	Alpha Flows
		DOS Attack
	Rooted Tree	Flash Crowd
		Port Scan
Ant-Inspired Algorithm	BiPartite	Network Scan
		Outage Event
	Others	Point to Multipoint
		Targeted Attack (e.g. Phishing)

# Theory Technical Challenges

- Natural Selection doesn't produce **the best** solution to a problem, it simply eliminates **failed** solutions – after borrowing initial inspiration, we will need to consider how the system could be improved upon for our purposes
- Bridging the “language gap” between theoretical biology and applied computer science
- (not within current scope, but to be considered) Robustness of system to design of next-generation threats that would specifically attempt to evade distributed detection.

# Implementation System Challenges

- Mapping biological concepts to computer networks
- Choosing the right kinds of anomalies
- Creating sound experiments on DETER
- Validating our results

# Milestones & Deliverables

Milestones & Deliverables	Team	Date
Fully Documented Design for Ant and Bee Algorithms	RU	5/13
Fully Documented Network Design and Prototype	Sub	
Completed Software Environment for Testing	RU	5/13
Completed Software-Based Performance Evaluation for One Algorithm on Specific Anomaly Types (preparation of publications and presentations)	RU	6/13
Papers on Experiment Design and Analysis of Results	Sub	
Completed Software-Based Performance Evaluation for Other Algorithm on Specific Anomaly Types (preparation of publications and presentations)	RU	12/13
Fully Documented Network Design and Prototype and Experimentation Results (preparation of publications and presentations)	Sub	

# Milestones & Deliverables cont.

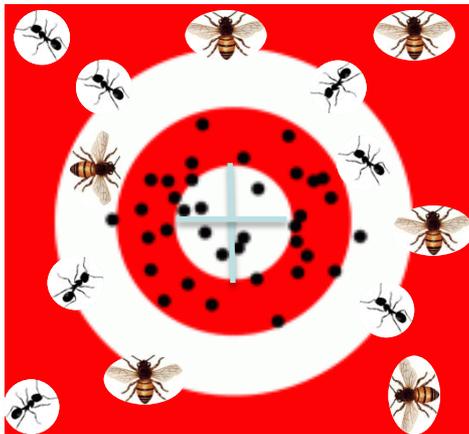
Milestones & Deliverables	Team	Date
Papers on Experiment Design and Analysis of Results (preparation of publications and presentations)	Sub	
Completed Software-Based Performance Evaluation for One Algorithm on Novel Anomaly Types (preparation of publications and presentations)	RU	5/14
Design General Method for Algorithm Performance Prediction	RU	10/14
Large-Scale Experimentation Phase 1	Sub	6/15
Large-Scale Experimentation Phase 2	Sub	10/15
Revise/Tailor Most Promising Algorithm for Large-Scale Testbed	RU	10/15

# Technology Transfer Plan

- Build working POC on ONE Blades
- Work with corporate environments to detect real-world anomalies quickly
- Work with incident investigators to determine detection rate and false positive rates

# Quad Chart

## *Distributed Intelligence*



finds the *elusive* adversary

### **Operational Capability**

- Performance targets:* Basic principles in 12 mo; Proof of concept in 24 mo; (Option) field testing and tech transfer in 36 mo.
- Quantify performance for key parameters:* Key performance measures derived in Year 1 will be used to evaluate effectiveness for appropriate botnet detection scenarios
- Cost of ownership:* None. Project results will be in public domain
- Address how the proposed development addresses the goals in the BAA.* Provides scalable distributed intelligence for detecting hard-to-find malware-induced behavior; leverages biological understanding of bees and ants to design communication protocols; results in significant tech transfer

### **Proposed Technical Approach**

- Addressing goals in the BAA:* Models biological systems for new methods for cyber-health plus technology transfer.
- Base Period tasks:* Define distributed detection algorithms; Implement and test software simulations to test algorithms on simple network topologies; Build networking substrate; Test and evaluate anomaly detection performance on a diversity of anomalies.
- Current status:* The biological phenomena have already been studied by the proposer. Proposed work will marry this with cyber security.
- Describe any actions done to date.* None. This is a fresh proposal.
- Describe any related ongoing effort by the offeror:* Distributed correlation capability is on a short term list in at least one HP security product unit and in HP networking.

### **Schedule, Cost, Deliverables, Contact Info**

*Milestones:* Biology-based detection algorithms designed and evaluated December 2012; ProCurve Networking prototype delivered December 2013; Tech transfer December 2014

*Period of performance:* 3 years

*Deliverables:* Application of basic principles of bio-inspired distributed detection;

Enhanced network switches with detection;

Decentralized switch protocols for data sharing; Consolidated prototype; Tech transfer

*Corporate Information:*  
 Sarah Dumais, Rutgers University, 3 Rutgers Plaza, New Brunswick, NJ 08901, phone: 732-932-0150 x 2107, fax: 732-932-0162, email: dumais@grants.rutgers.edu