Self-organized Collaboration of Distributed IDS Sensors

Karel Bartos[1] and Martin Rehak[1,2] and Michal Svoboda[2]

[1] Faculty of Electrical Engineering
Czech Technical University in Prague

[2] Cognitive Security, s.r.o., Prague

DIMVA 2012 July 27 2012

Network Security – Motivation

- **Advanced Persistent Threats**
  - Strategically motivated
  - Targeted (single/few targets)
- **Threats**
  - Sophisticated industrial espionage
  - Organized crime – credit card fraud, banking attacks, spam
- **Challenges**:
  - High traffic speeds
  - High number of increasingly sophisticated, evasive attacks

agent technology center

All Industry Sectors at Risk

U.S. Federal Government	6
U.S. State Government	5
U.S. County Government	3
Canadian Government	2
South Korean Government	1
Vietnam Government	1
Taiwan Government	1

Construction/ Heavy Industry	3
Steel Industry	1
Energy	1
Solar Power	1
U.S. Government Contractor	1
United Nations	1
Indian Government	1

Electronics Industry	3
Computer Security	2
Information Technology	2
Satellite Communica-tions	2
News Media	2
Information Services	1
Communica-tions	1

Defense Contractor	13

Real Estate	2
Accounting Industry	2
Agriculture	1
Insurance	1

International Sports	5
Economics/ Trade	2
Think Tanks	2
International/ Government/ Economics/ Trade	1
Political Non-profit	1
U.S. National Security Non-profit	1

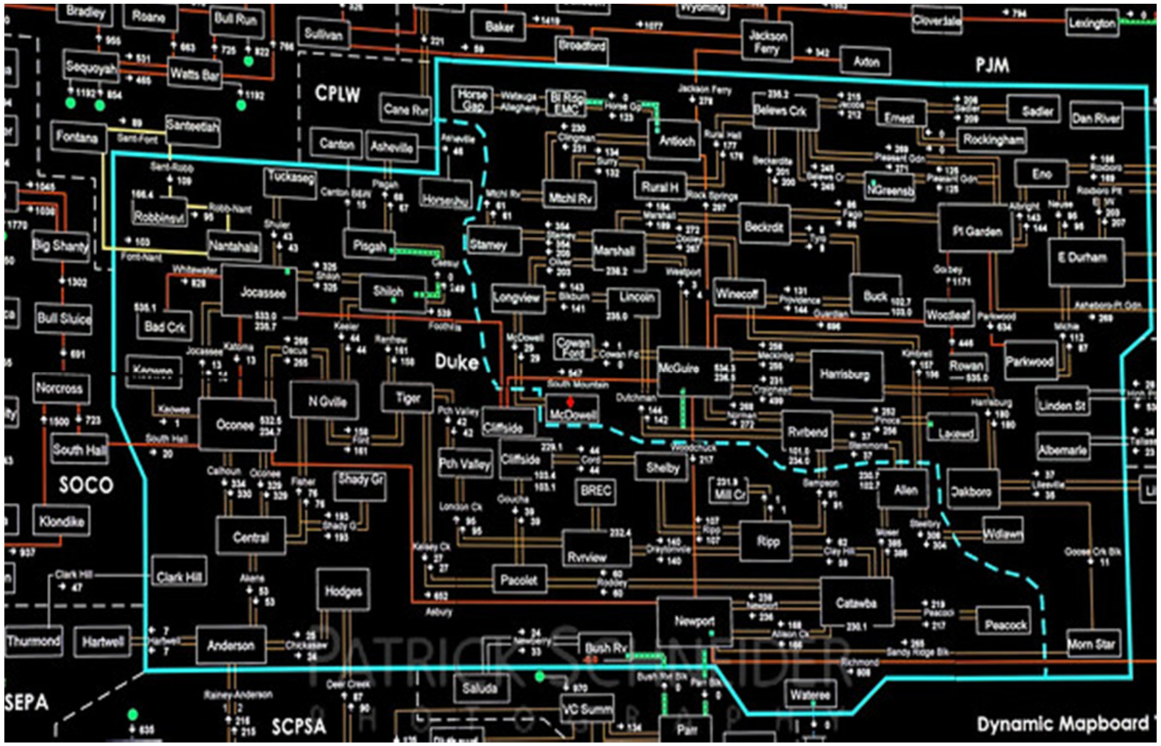
"…every company in every conceivable industry with significant size & valuable intellectual property & trade secrets has been compromised (or will be shortly)…" - McAfee

agent technology center

McAfee - Revealed, Operation Shady RAT

Our Goal

- **Use a Collaboration of Multiple Heterogeneous Detectors to create Network Security Awareness**



agent technology center

Intrusion Detection

- **Intrusion Detection Systems**
  – Deployed on key points of the network infrastructures
  – Detects malicious network/host behavior
- **Approaches**
  – Host based vs. Network based
  – Anomaly detection vs. Signature matching
  – Multi-algorithm systems
- **Problem:** Stand-alone IDS is not very effective on
  – Cooperative attacks
  – Large variability of malicious behavior

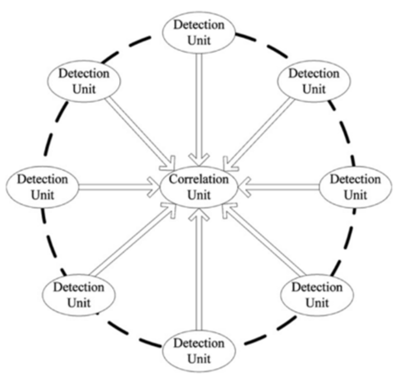
Current Solution? Alert Correlation

- **IDEA: Data fusion of results from more detectors**
- **GOAL: Create global full scale conclusions**
  - Fusion of raw input data or low-level alerts
  - Increase the level of abstraction
  - Reveal more complex attacks scenarios
  - Find prerequisites and consequences

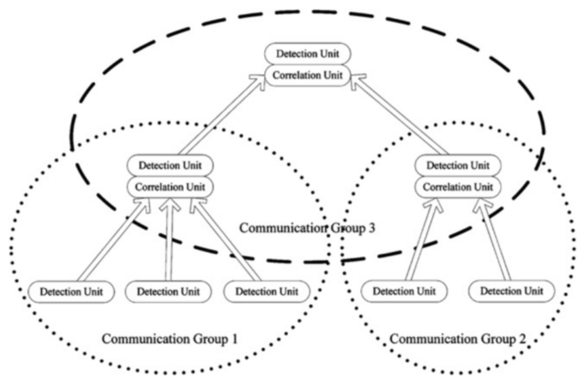
Alert Correlation

- **Architectures**

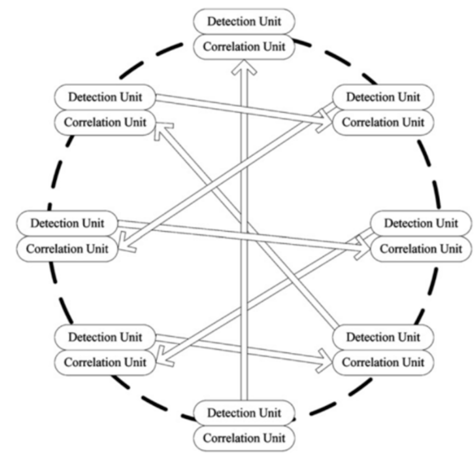
Centralized



Hierarchical



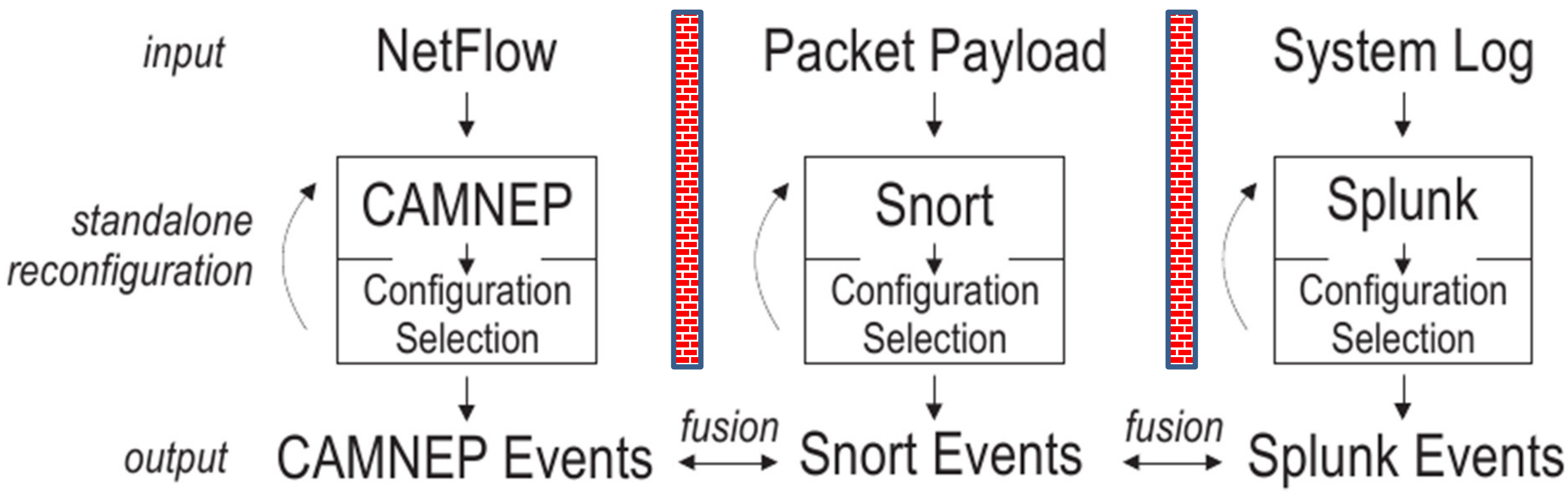
Fully-distributed



agent technology center

Example of Current Architecture

– All detectors work in a **stand-alone** architecture
– More sophisticated detectors can reconfigure based on **local** observations



agent
technology
center

Alert Correlation

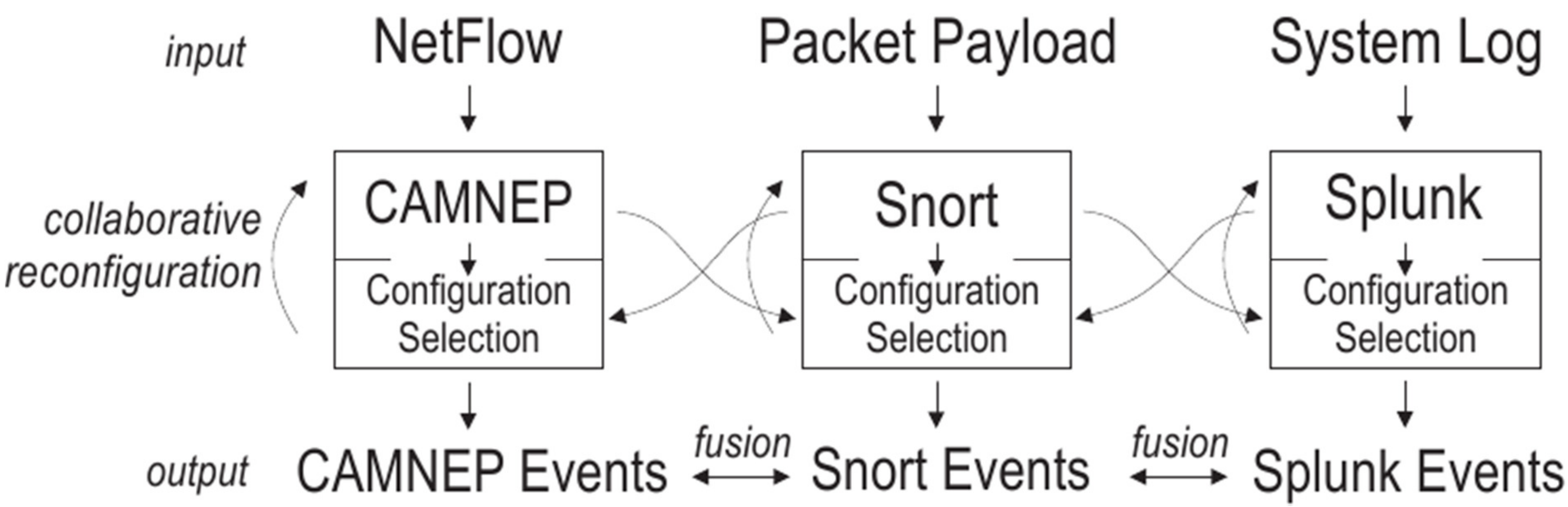
- **Collects results from more detectors to provide better overall results**

- **WEAKNESSES:**
- **It does not provide any feedback to the detectors**
  - Detectors are not aware of the performance of other detectors
  - Detectors require initial (manual) configuration/tuning
- **It does not improve the performance of detectors**

Our Approach

– All detectors work in a **fully distributed** and **collaborative** architecture

– More sophisticated detectors can improve based on observations from other detectors



Assumptions and Requirements

- **Communication**
  - All-to-All, fully distributed
- **Reconfiguration**
  - At least some detectors are able to change their internal states according to the observations
- **Security**
  - Detectors do not provide information about their internal states
- **Strategic Deployment**
  - Detectors are deployed in various parts of the monitored network; network traffic should overlap

Why to communicate and share results?

- **Large variability of network attacks and threats**
  - No single detector is able to detect all intrusions
- **To detect more intrusions, we need more detectors**
  - More detection methods, various locations
- **Many detectors report a lot of same intrusions**
  - They make similar conclusions and mistakes

agent
technology
center

Why to communicate and share results?

- **Large variability of network attacks and threats**
  - No single detector is able to detect all intrusions
- **To detect more intrusions, we need more detectors**
  - More detection methods, various locations
- **Many detectors report a lot of same intrusions**
  - They make similar conclusions and mistakes

Q: Is it a good thing?

agent technology center

Why to communicate and share results?

- **Large variability of network attacks and threats**
  - No single detector is able to detect all intrusions
- **To detect more intrusions, we need more detectors**
  - More detection methods, various locations
- **Many detectors report a lot of same intrusions**
  - They make similar conclusions and mistakes

Q: Is it a good thing?

- For traditional alert correlation: **YES** (FP reduction)

agent technology center

Why to communicate and share results?

- **Large variability of network attacks and threats**
- **To detect more intrusions, we need more detectors**
- **Many detectors report a lot of same intrusions**

Q: Is it a good thing?

– For traditional alert correlation: **YES** (FP reduction)

Q: Why the detectors generate a lot of FP?

agent technology center

Why to communicate and share results?

- **Large variability of network attacks and threats**
- **To detect more intrusions, we need more detectors**
- **Many detectors report a lot of same intrusions**

Q: Is it a good thing?

– For traditional alert correlation: **YES** (FP reduction)

Q: Why the detectors generate a lot of FP?

A: Because they: - want to be universal

- want to generate a lot of TP

agent technology center

Why to communicate and share results?

- **Large variability of network attacks and threats**
- **To detect more intrusions, we need more detectors**
- **Many detectors report a lot of same intrusions**

Q: Is it a good thing?

- For traditional alert correlation: **YES** (FP reduction)
- For our approach: **NO** (**specialization**)

Specialization

- **IDEA: Detectors communicate in order to be special**
- **Each detector wants:** (specialization allows)
  - to detect unique intrusions → *essential*
  - to minimize the amount of FP → *effective*
- **Each detector does not want:** (specialization prevents)
  - to waste resources on already detected intrusions
- **Specialization in collaboration**
  - Maximizes the overall detection potential of the system

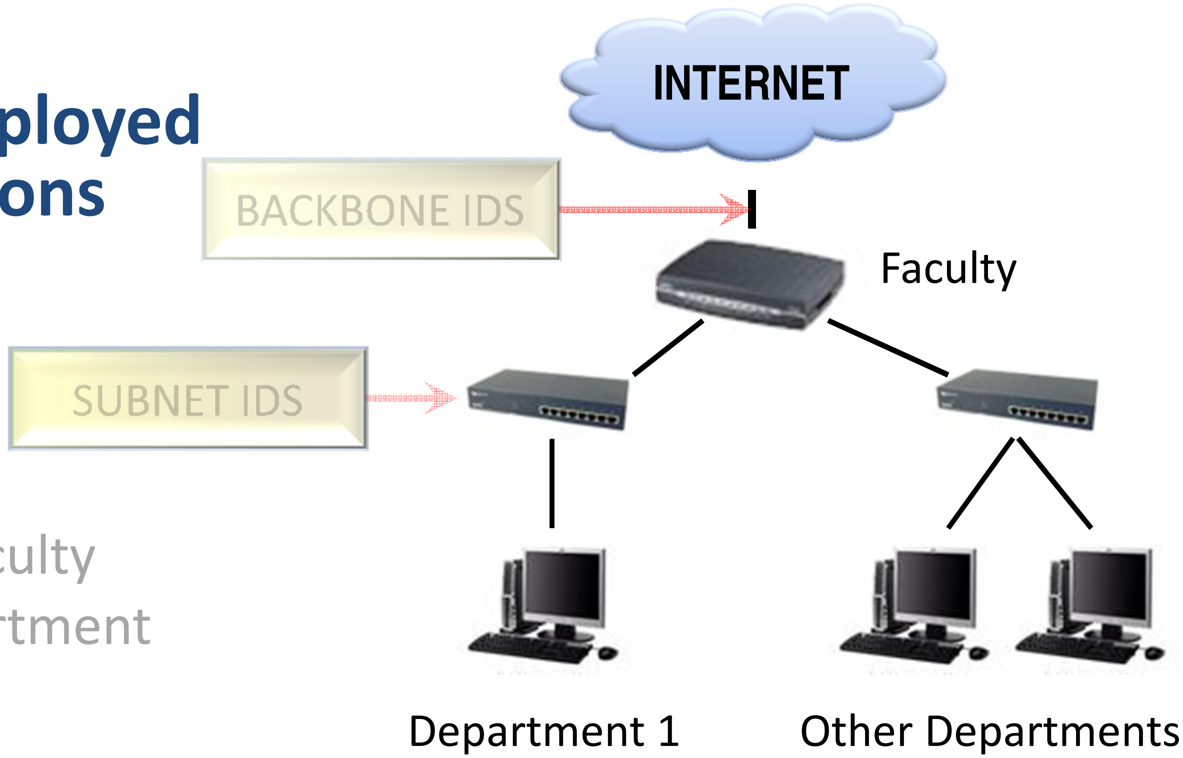
agent technology center

Proposed Collaboration Model

- **Set of feedback functions**
  - Computes the specialization of each detector
  - f: E_local × E_remote → **R**
- **Set of configuration states**
  - Defines the behavior of each detector
- **Solution Concept / Algorithm / Strategies**
  - Feedback – reconfiguration mapping
  - Suitable for dynamic network environments

Experimental Evaluation - Setup

- **2 network IDS deployed in different locations of our University network**

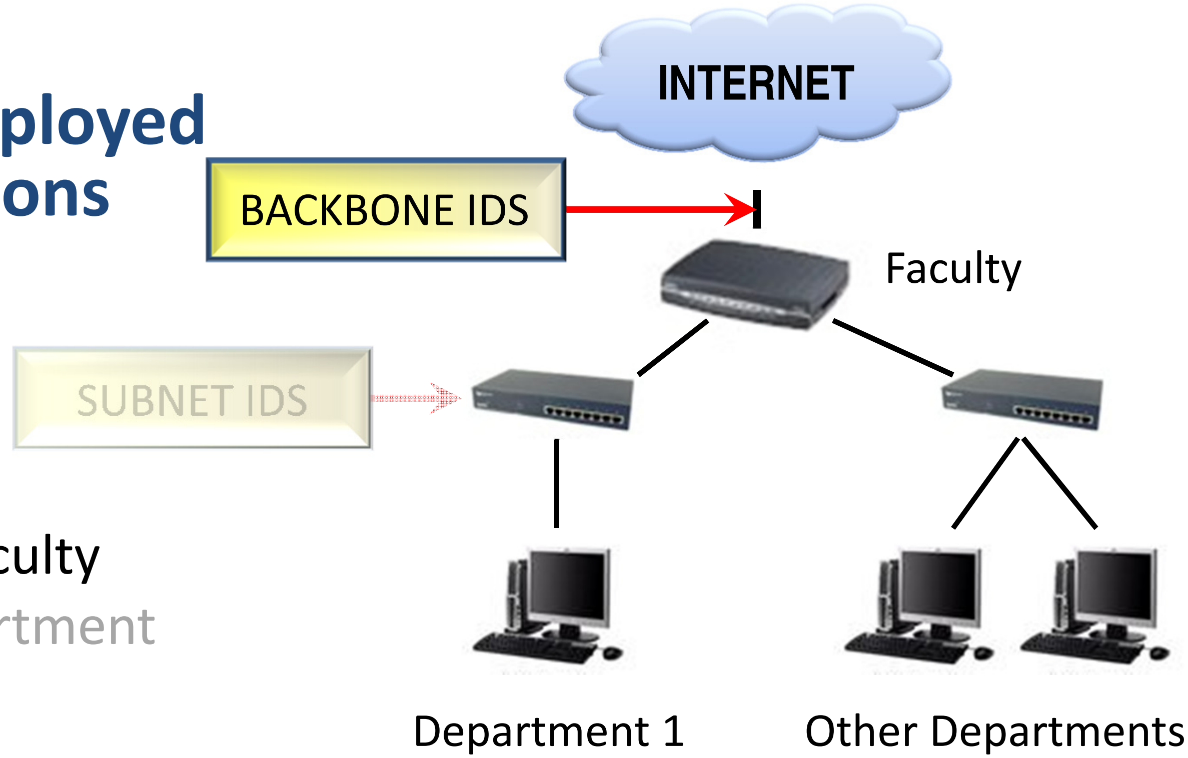


  - Backbone IDS – Faculty
  - Subnet IDS – Department
  - 10 hours of network traffic (NetFlow)
  - Including samples of malware behavior

agent technology center

Experimental Evaluation - Setup

- **2 network IDS deployed in different locations of our University network**

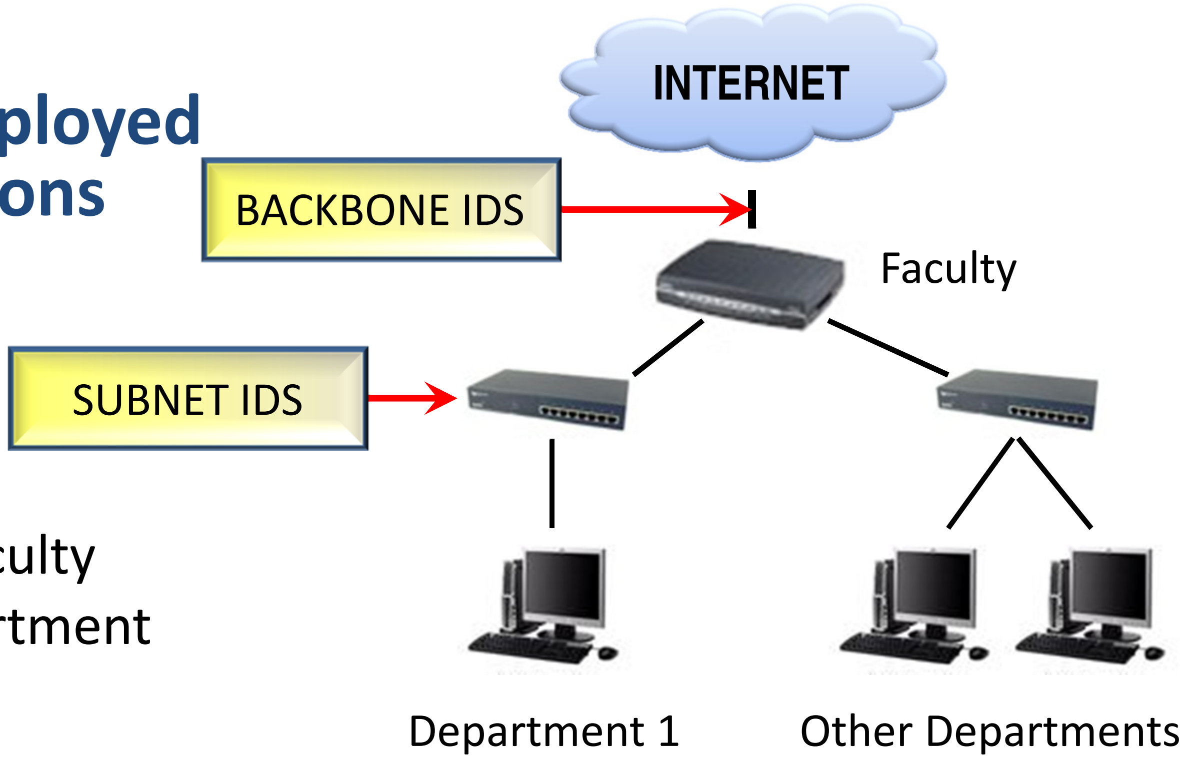


  - Backbone IDS – Faculty
  - Subnet IDS – Department
  - 10 hours of network traffic (NetFlow)
  - Including samples of malware behavior

agent technology center

Experimental Evaluation - Setup

- **2 network IDS deployed in different locations of our University network**

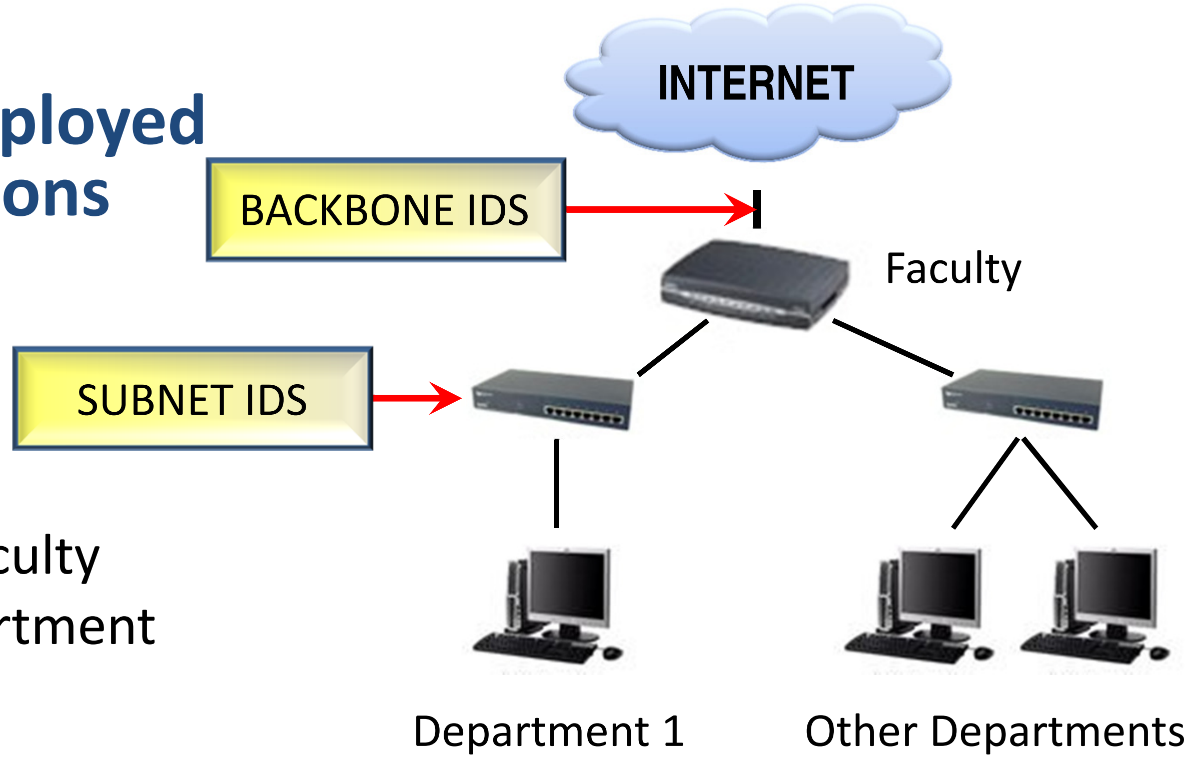


  - Backbone IDS – Faculty
  - Subnet IDS – Department
  - 10 hours of network traffic (NetFlow)
  - Including samples of malware behavior

agent technology center

Experimental Evaluation - Setup

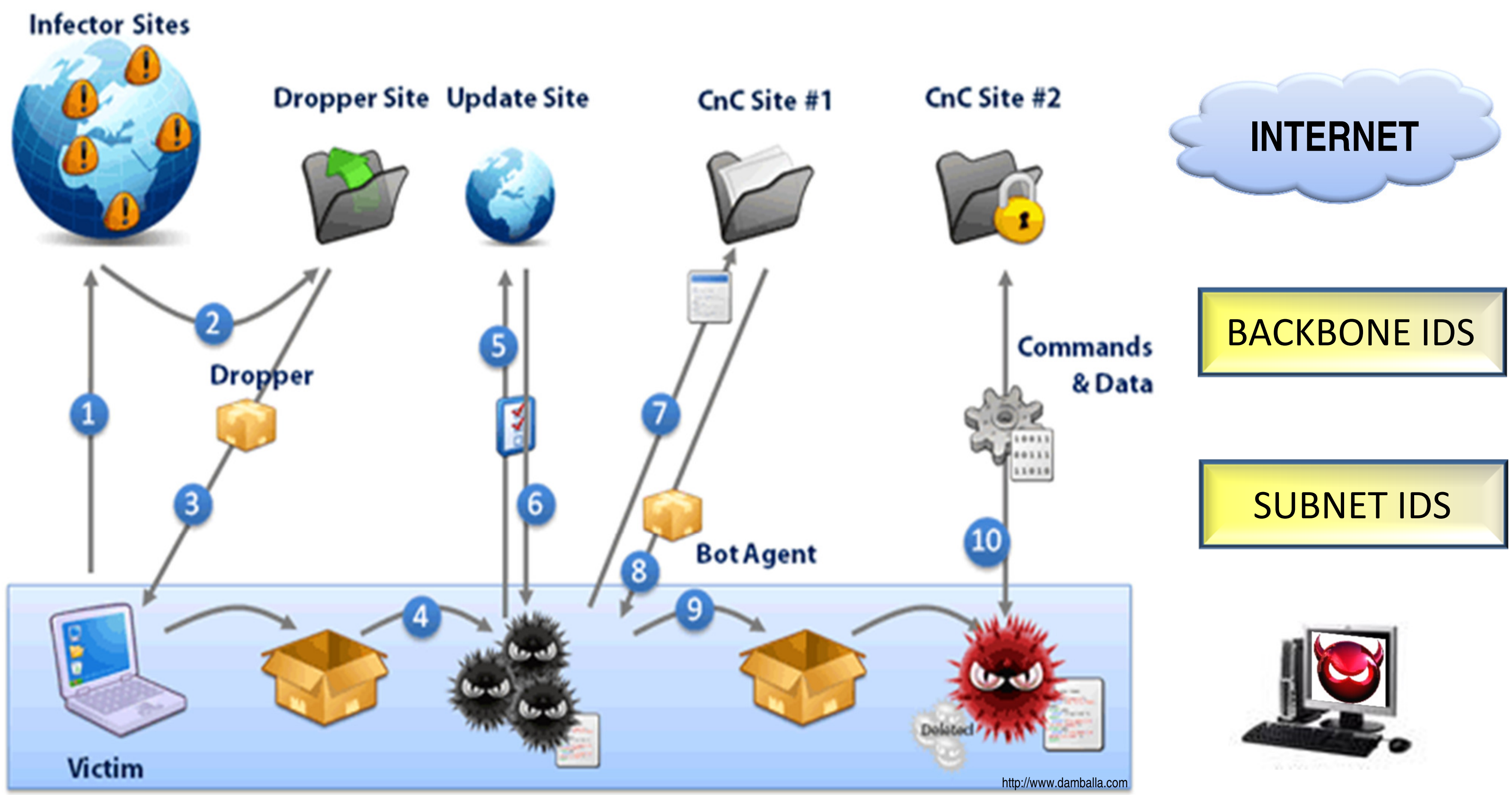
- **2 network IDS deployed in different locations of our University network**



  - Backbone IDS – Faculty
  - Subnet IDS – Department
  - 10 hours of network traffic (NetFlow)
  - Including samples of malware behavior

agent technology center

Experimental Evaluation - Malware



agent technology center

Experimental Evaluation - Model

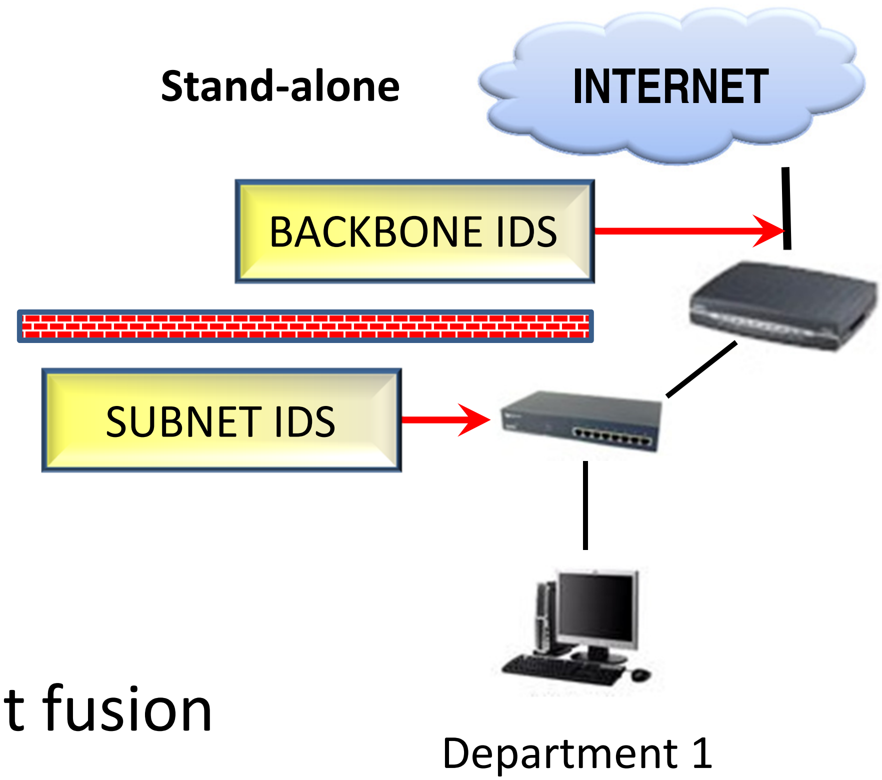
- **Feedback function is defined as**
  - Uniqueness of generated events
  - Number of alerts that I detected and others did not
- **Set of configuration states**
  - Each detector consists of several detection methods
  - Several opinions have to be aggregated = parameter

  - State = aggregation function within each IDS

agent technology center

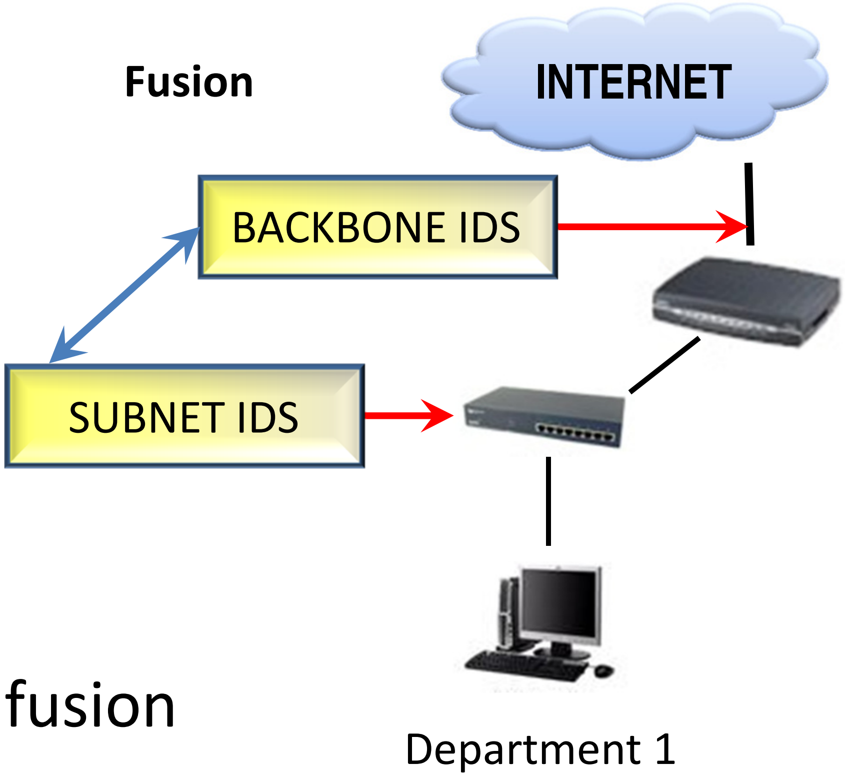
Experimental Evaluation - Strategies

- **Stand-alone**
  - No feedback, No fusion
- **Fusion only**
  - Detectors are connected and exchange their results
- **Fusion + Feedback**
  - Distributed feedback, Event fusion
  - Encourages specialization



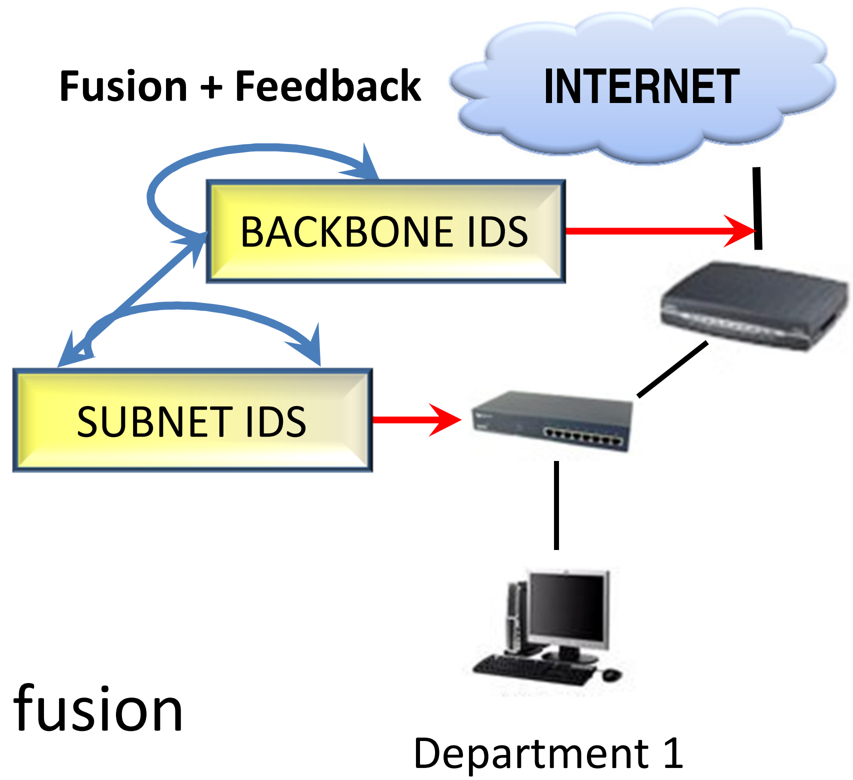
Experimental Evaluation - Strategies

- **Stand-alone**
  - No feedback, No fusion
- **Fusion only**
  - Detectors are connected and exchange their results
- **Fusion + Feedback**
  - Distributed feedback, Event fusion
  - Encourages specialization



Experimental Evaluation - Strategies

- **Stand-alone**
  - No feedback, No fusion
- **Fusion only**
  - Detectors are connected and exchange their results
- **Fusion + Feedback**
  - Distributed feedback, Event fusion
  - Encourages specialization

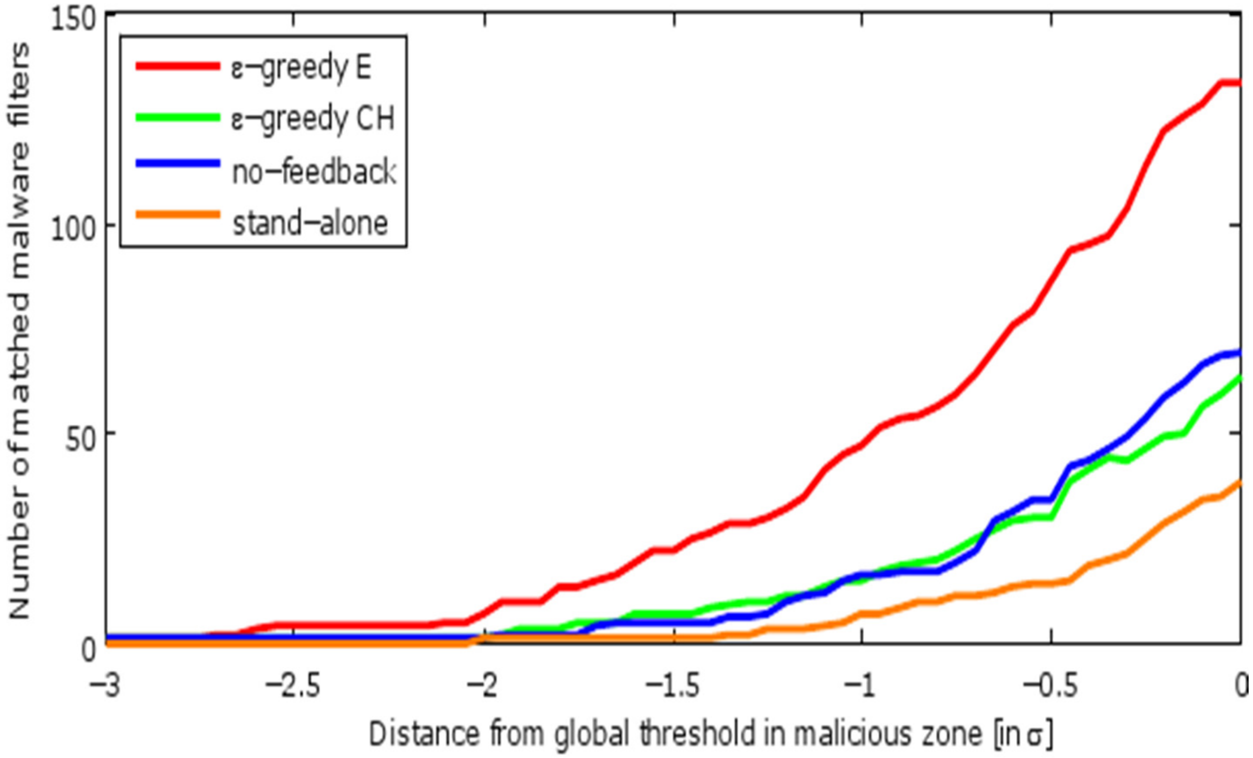


FIRE Epsilon-greedy Adaptation

- **Model consists of configuration states and their uniqueness values (weighted 5 past values)**
- **Algorithm**
  - Detectors exchange events
  - Compute uniqueness of last used configuration
  - Update last 5 uniqueness values for last used configuration
  - With probability p:
    - $p \geq \varepsilon$ select most unique configuration
    - $p < \varepsilon$ select random configuration

Experimental Evaluation - Results

- **Subnet location – # of detected malware samples**



132 Feedback + Fusion

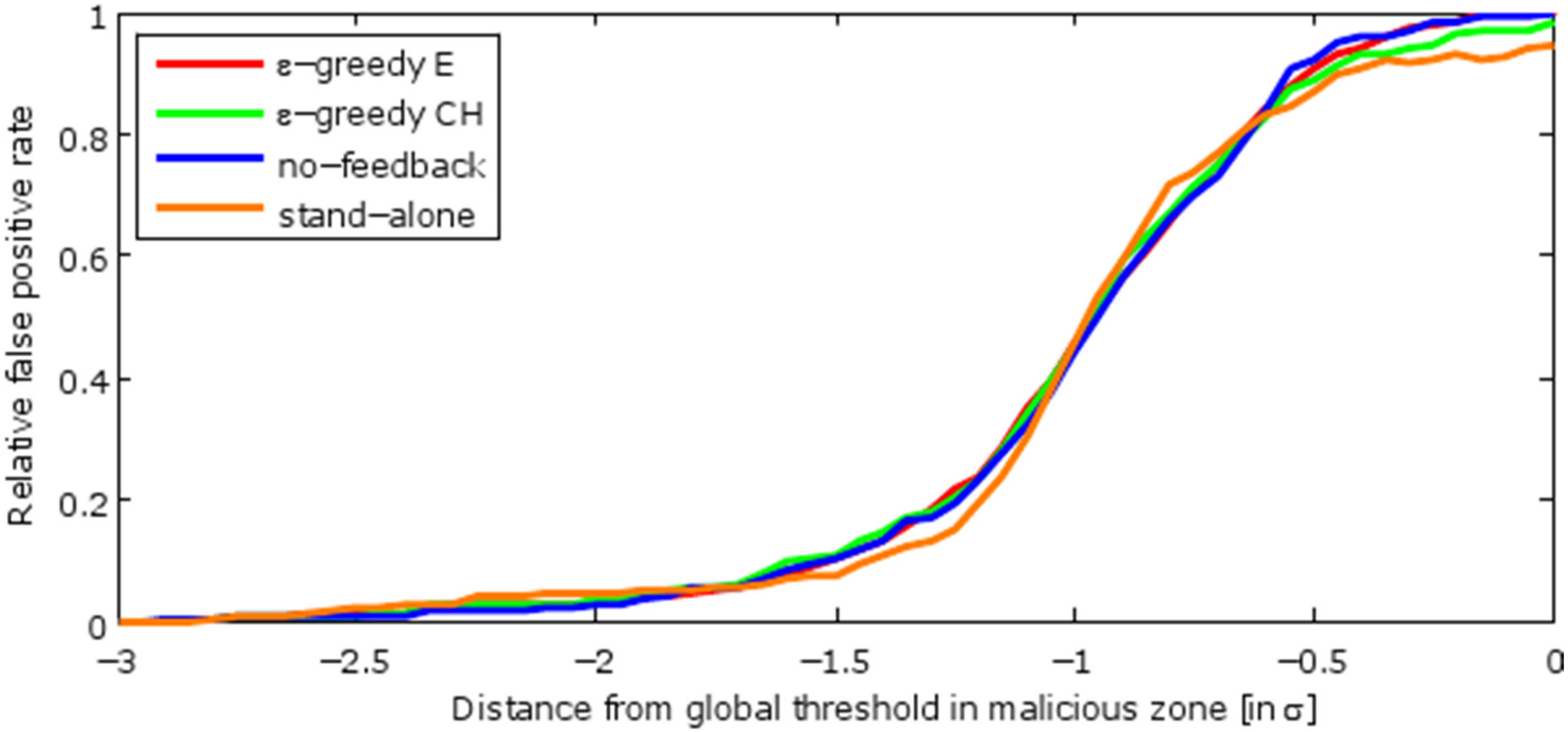
71 Fusion only

38 Stand-alone

agent technology center

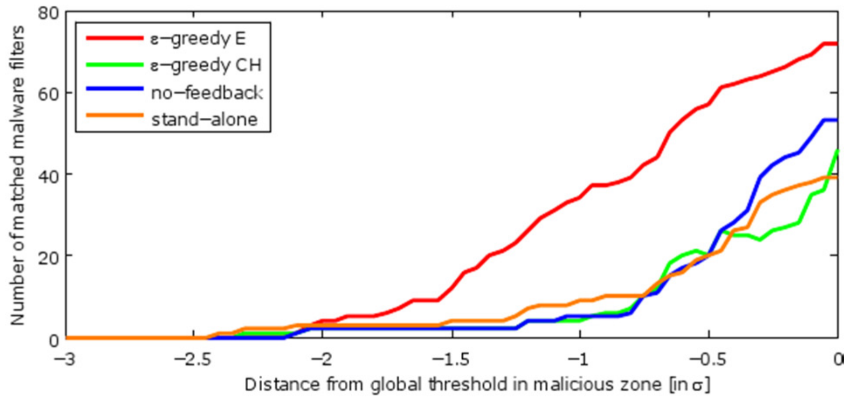
Experimental Evaluation - Results

- **Subnet location – relative false positive rate**



Experimental Evaluation - Results

- **Backbone location – # of detected malware samples**

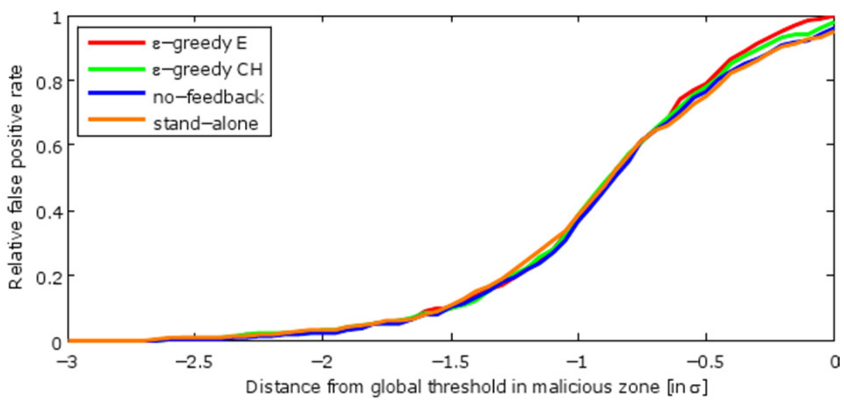


72 Feedback + Fusion

53 Fusion only

39 Stand-alone

- **Backbone location – relative false positive rate**



agent technology center

Conclusion

- **Distributed collaboration of heterogeneous detectors**
- **Extends overall detection potential of the system by mutual specialization of the detectors**
- **Future Work:**
  - Other strategy selection techniques
  - More extensive experimental evaluation

Thank You

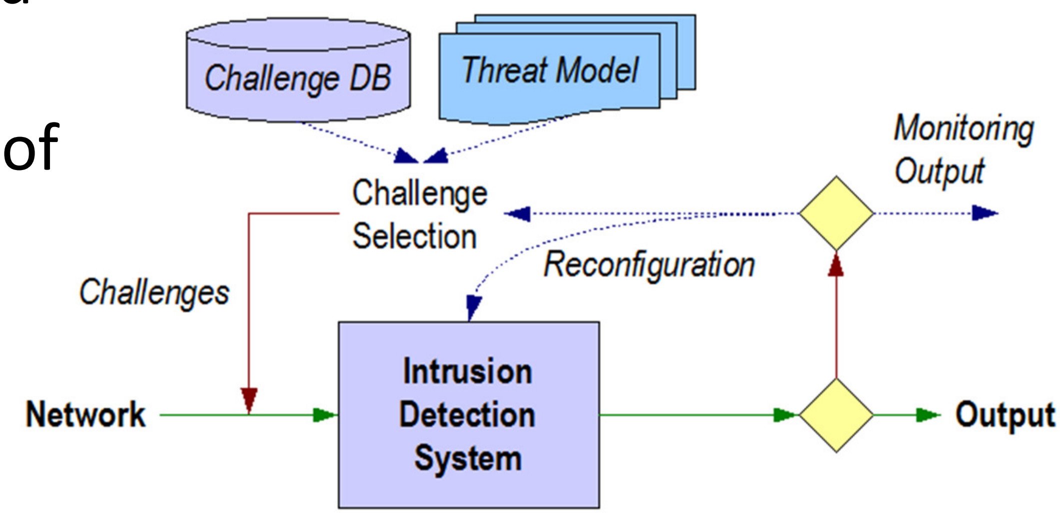
Questions?

Thank You

Questions?

Local Self-adaptation

- Unlabeled background input data
- Insertion of small set of challenges
  - Legitimate
  - Malicious
- Response evaluation
- Problems: Noise, challenge non-uniformity, distribution, system compromise



Challenge Insertion Control