**DISTDET: A Cost-Effective Distributed Cyber Threat Detection System**

Xusheng Xiao, Associate Professor, School of Computing and Augmented Intelligence, Arizona State University

---

**Advanced Persistent Threat (APT) Attack**

APT attacks have plagued many well-protected businesses

- Advanced: sophisticated techniques exploiting multiple vulnerabilities
- Persistent: continuously monitoring and stealing data from target
- Threat: strong economical or political motives

---

**Ubiquitous System Monitoring**

- Build a provenance graph based on system events collected from system kernels, describing operations of system entities (e.g., process read/write files).
- Contextual information in the provenance graph is effective in revealing advanced attack tactics.

---

**False Alarm Filtering**

- Observation: "the alarms representing the same behaviors will be repetitively reported over a period of time"
- Alarm Deduplication: Alarms with the same events in a time window are aggregated into one alarm
- Observation: "many false alarms are related to the benign behaviors triggered by semantically similar commands"
- ASG Semantic Aggregation: ASGs with similar commands are aggregated into one ASG
- Observation: "the contexts for these alarms are generally known to represent benign behaviors"
- ASG Prioritization: Compute anomaly scores of aggregated ASGs based on rareness (frequency) and filter those with low anomaly scores

---

**Global Model Derivation**

**Observation:** "local models can easily lead to false alarms in detection"
- Cluster the host models based on the services provided by the hosts
- Extract the list of service-specific processes from each host model
- Compute the word embeddings of the extracted processes’ names
- Use k-means algorithm to cluster the host models
- Merge the behaviors of the common processes in the same cluster

---

**DISTDET System**

The first cost-effective detection system that synergically combines distributed computing, anomaly detection, and false alarm filtering techniques for detecting and investigating advanced cyber attacks

- Lightweight Client-Side Detection
  - Shift part of the attack detection to the clients and transmit only summary graphs that represent potential attacks to the server.
- Unique Properties of False Alarms
  - False alarms typically possess some unique properties: (1) the alarms representing the same behaviors will be repetitively reported over a period of time; (2) many false alarms are related to the benign behaviors triggered by semantically similar commands; (3) the contexts for these alarms are generally known to represent benign behaviors.
- Global View of Service Behaviors
  - A global model built in the server can observe the behaviors in all the phases and can complement the missing observations in the local models.

---

**Host-based Anomaly Detection – HST and ASG**

- Hierarchical System Event Tree (HST) is a compact index that categorizes auditing events based on their properties using a multi-layer tree. It is built based on training events. Any event not observed in the built HST will generate an alarm.
- An Alarm Summary Graph (ASG) is a summary graph that includes the process p that initiates the suspicious behavior reported in an alarm, and the events initiated by p’s ancestor processes and descendant processes.

---

**Evaluation Summary**

- Reduce the host cost (the expense of securing a single host) from 3.4 USD to 0.061 USD (56 × reduction)
- Outperform the state-of-the-art approaches
- Achieve a F1 of 0.98 for the industry arena and DARPA TC datasets
- Achieve a F1 of 0.89 for the public arena dataset
- DISTDET reduces the false alarms from 230 alarms/host/day to 0.71 alarms/host/day, saving 99.69% of the required inspection efforts.
- DISTDET found 900+ real attacks during roughly 6 months and achieved better performance than other existing EDRs.