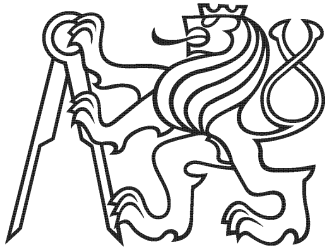


CAMNEP
Network Intrusion Detection System

Martin Rehak, Tomas Pevny, Michal Pechoucek,
Karel Bartos , Martin Grill, Jan Jusko, , Jan Stiborek, Jan Kohout
Ondrej Fikar, Marek Pytela
Czech Technical University in Prague

Network Intrusion Detection

- **Threats** in the computer networks
  - Sophisticated, possible government-backed attacks
    - industrial espionage, political attacks
  - Organized crime – credit card fraud, banking attacks, spam
- **Goal**: **Detection** of **computer attacks** by analyzing the **structure** of network traffic
  - privacy preserving
  - applicable on ciphered traffic (for financial/government sites)
  - similar to the analysis of phone bills
- **Challenges**:
  - High traffic speeds (millions of connections per second)
  - High number of increasingly sophisticated, evasive attacks

Anomaly Detection vs. Signatures

Signature matching

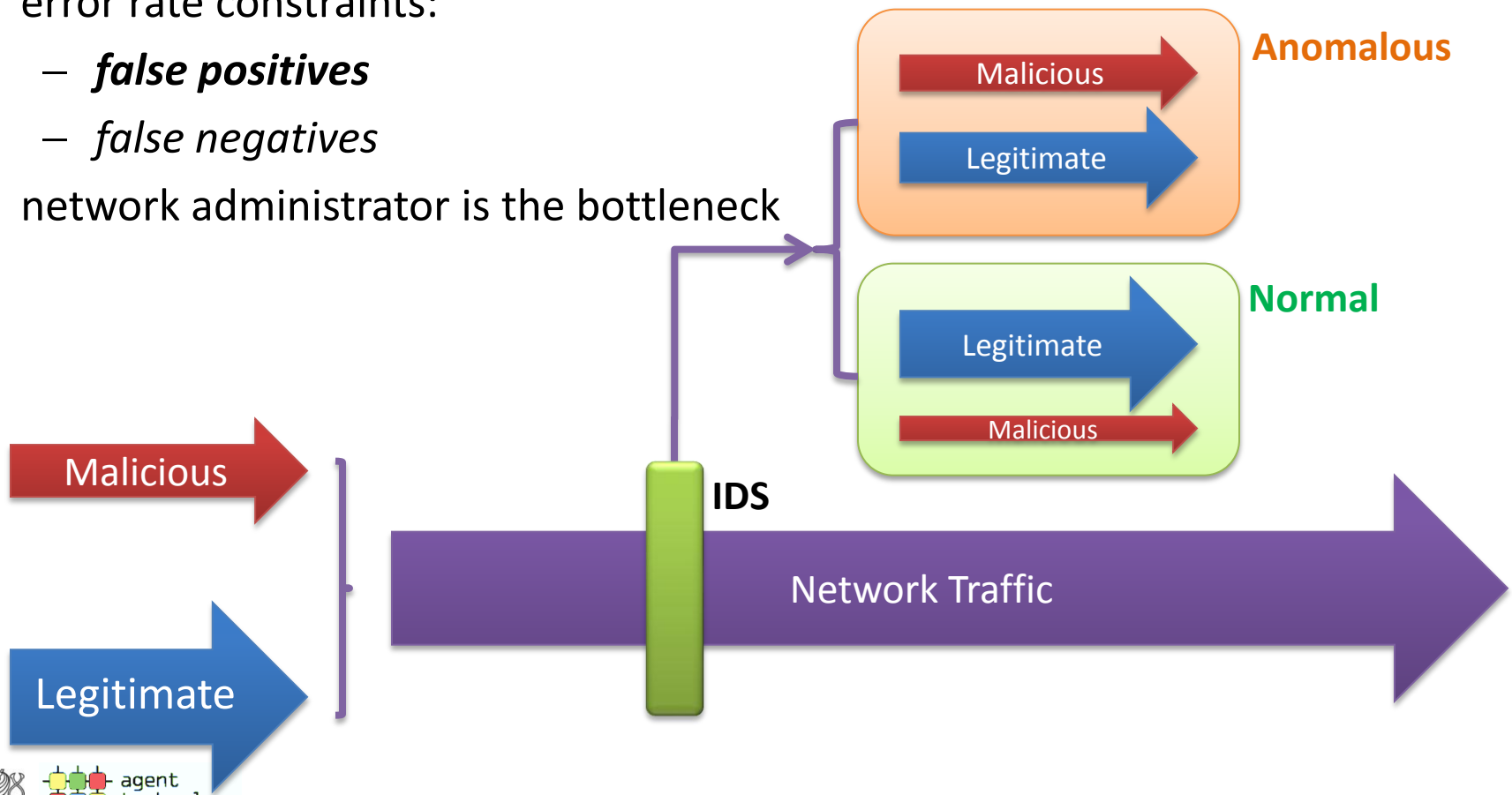
- Historically validated
- Widely deployed
- Verifiable & Stable
- Number of patterns
- Scaling
- Management
- New threats detection

Anomaly detection

- No patterns
- New threats detection
- Scaling
- **Error Rate/Sensitivity**
- **Verifiability**
- **Stability**
- **Management**

Anomaly detection

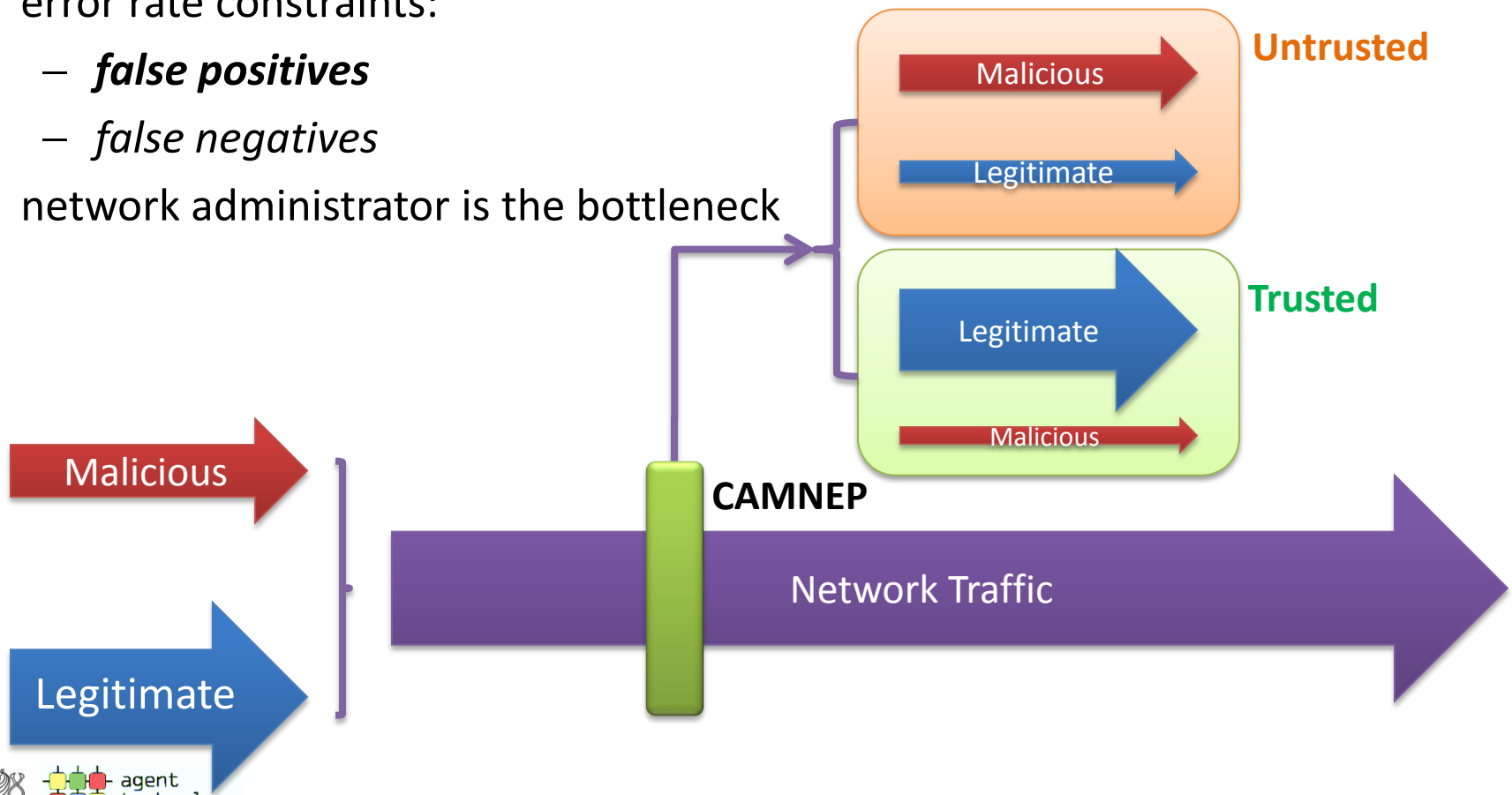
- real-time constraints: gigabit/sec, **2000** – **5000** flows/sec
- error rate constraints:
  - ***false positives***
  - *false negatives*
- network administrator is the bottleneck



agent technology center

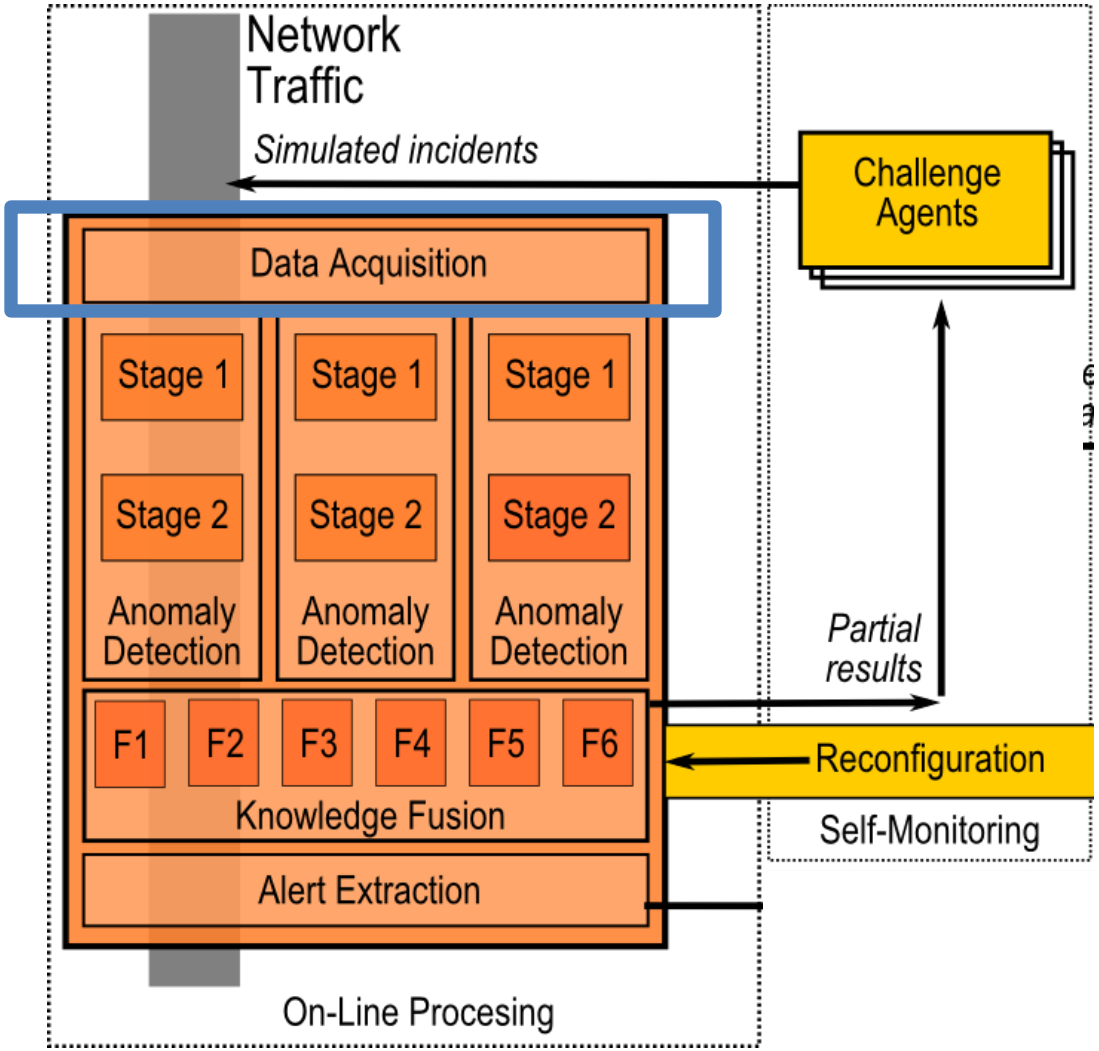
Anomaly detection

- real-time constraints: gigabit/sec, **2000** – **5000** flows/sec
- error rate constraints:
  - ***false positives***
  - *false negatives*
- network administrator is the bottleneck



agent technology center

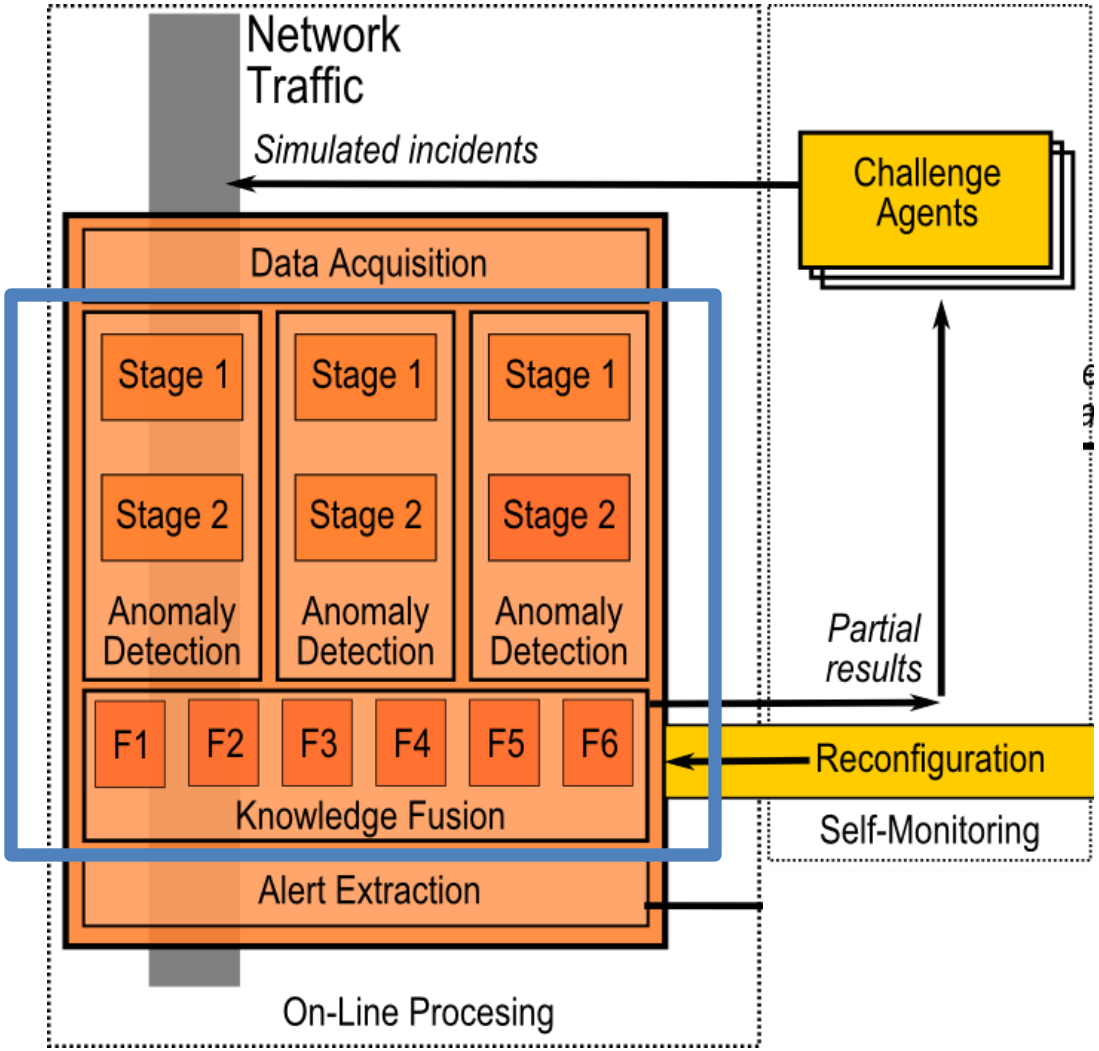
Architecture



- **Data Acquisition and Preprocessing Layer**
- Detection Layer
- Self-Monitoring Layer
- Alert Extraction Layer
- Analyst Interface

agent technology center

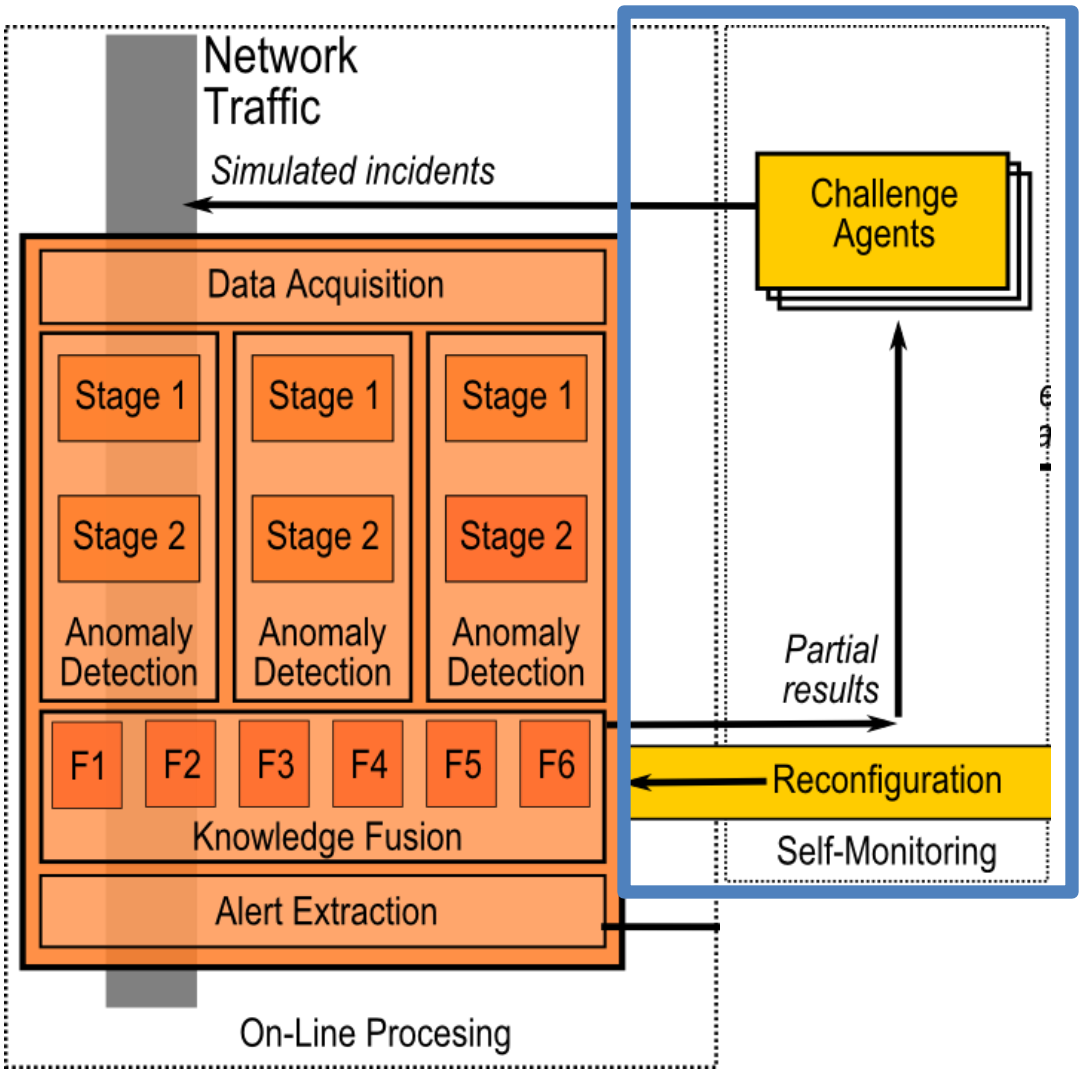
Architecture



- Data Acquisition and Preprocessing Layer
- **Detection Layer**
- Self-Monitoring Layer
- Alert Extraction Layer
- Analyst Interface

agent technology center

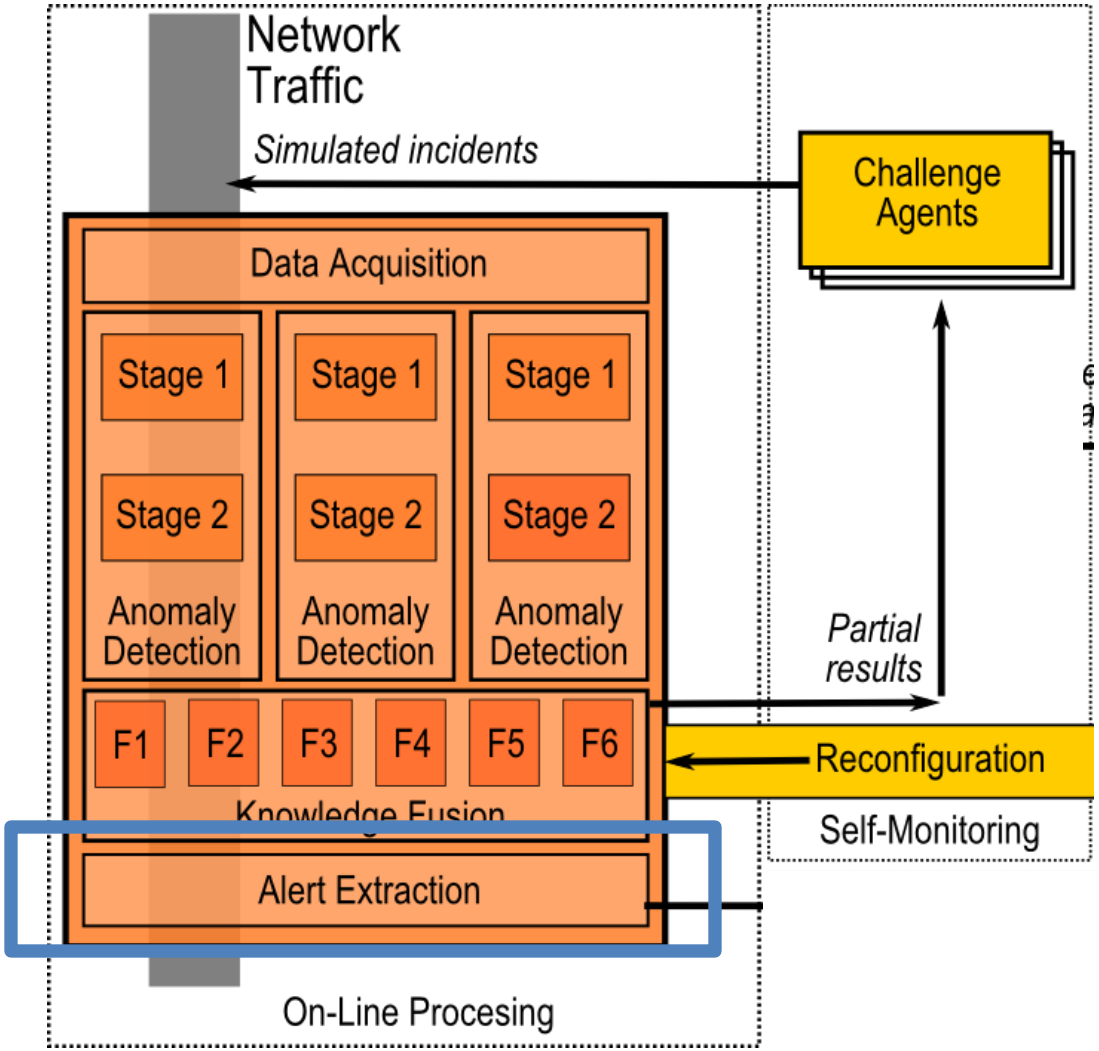
Architecture



- Data Acquisition and Preprocessing Layer
- Detection Layer
- **Self-Monitoring Layer**
- Alert Extraction Layer
- Analyst Interface

agent technology center

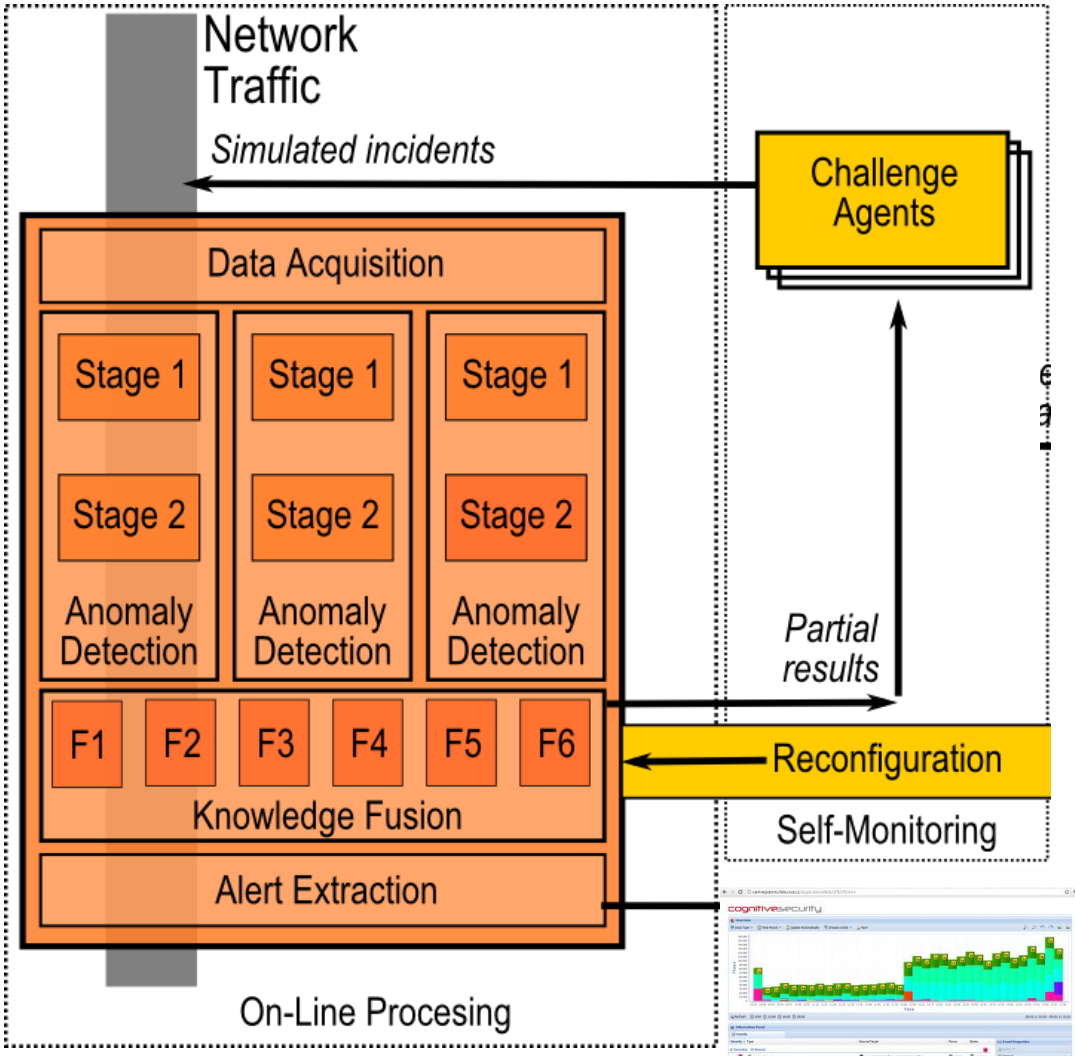
Architecture



- Data Acquisition and Preprocessing Layer
- Detection Layer
- Self-Monitoring Layer
- **Alert Extraction Layer**
- Analyst Interface

agent technology center

Architecture



- Data Acquisition and Preprocessing Layer
- Detection Layer
- Self-Monitoring Layer
- Alert Extraction Layer
- **Analyst Interface**

agent technology center

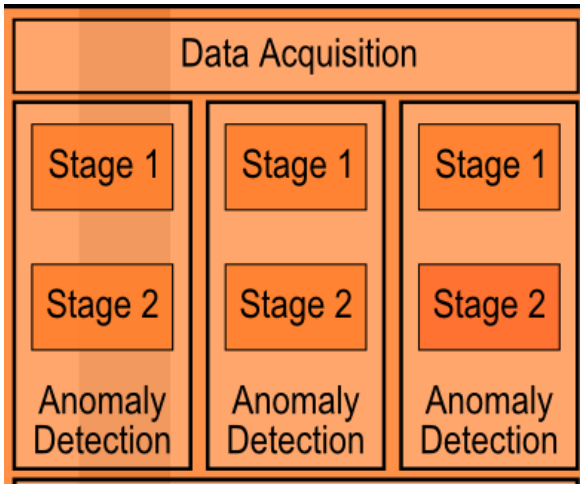
Detection Layer

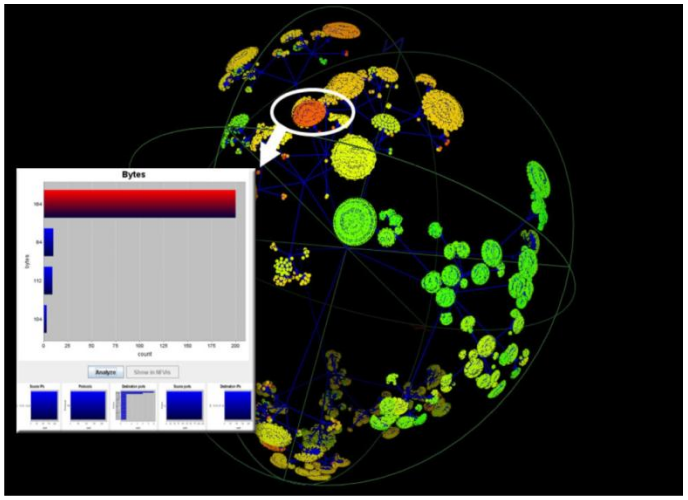
- **Anomaly Detection:** **Predicting current network behavior from the history and looking for deviations**

Detection Layer

- **Multi-Algorithm Anomaly Detection:**
- **Entropy modeling**
- **Trend modeling**
- **Volume modeling**
- **Principal components analysis**
- **Information-theoretical measures**
- **…**

Inside Modern NBA System



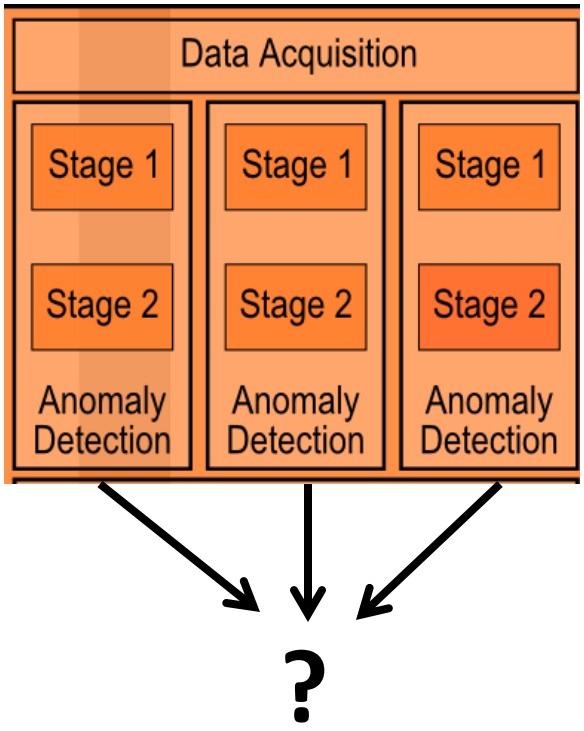


– **Trust Modeling: Synthesizing the Anomaly detection data across the algorithms and over time**

- **Reduction of false positives by:**
  Multi-source aggregation
  Historical experience aggregation

agent technology center

Key Problem n. 1



- **How to find optimal configuration = select aggregation function?**
- **We have many aggregation functions, but which is the best?**

Option 1: Offline Configuration

- Offline optimization of internal parameters
- Difficulties with training data
  - Expensive manual labeling
  - No fully labeled and representative dataset
  - Manually labeled data is biased
  - Legal issues with public sharing of data
- Offline configuration results can not capture the dynamic character of the network
  - M. Rehak, E. Staab, M. Pechoucek**, J. Stiborek**, M. Grill, and K. Bartos, "Dynamic information source selection for intrusion detection systems", International Foundation for Autonomous Agents and Multiagent Systems, 2009, pp. 1009–1016.

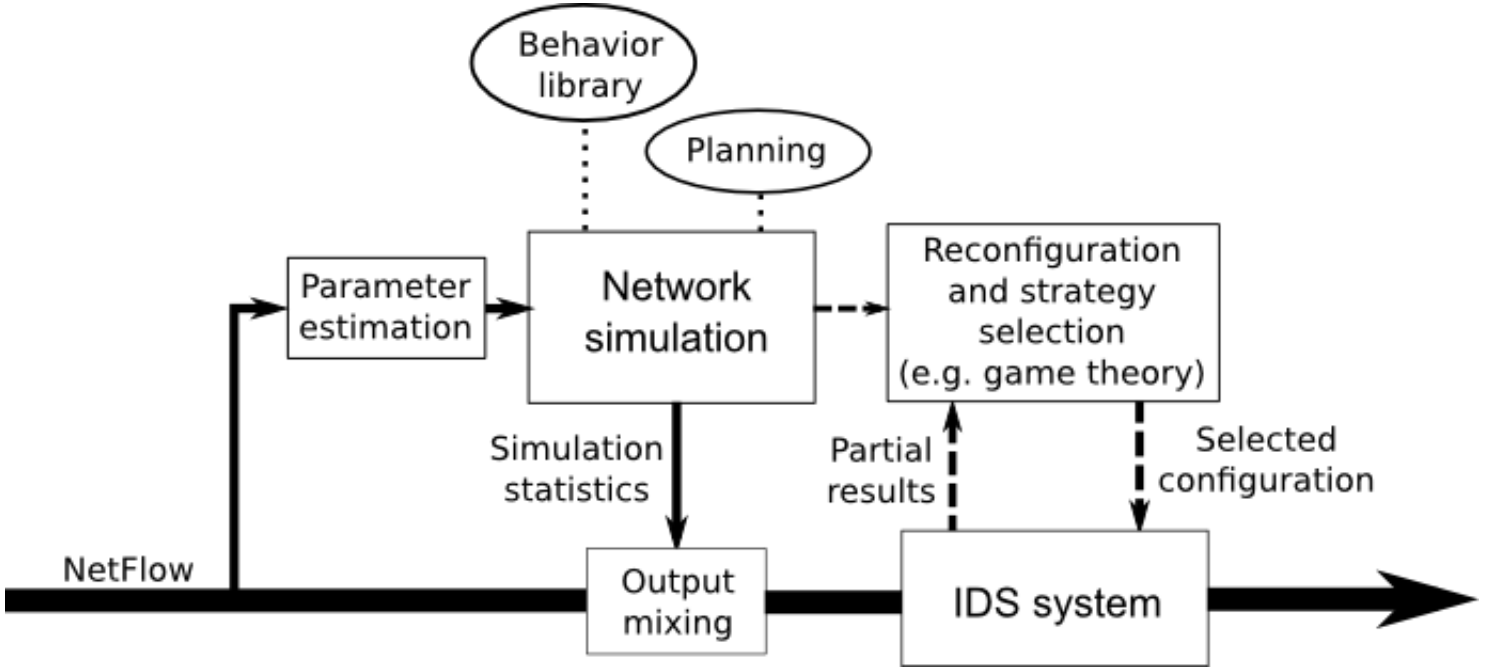
Option 2: Online Configuration

- Tunes parameters according to the current state of the specific network
  - Ground truth is undeterminable within the time constraints imposed
  - Manual, supervised approach is infeasible
- Malicious vs. Anomalous problem
  - Definition of malicious behavior depends on the specific network's security policy
  - Malicious vs. Anomalous fine-tuning problem

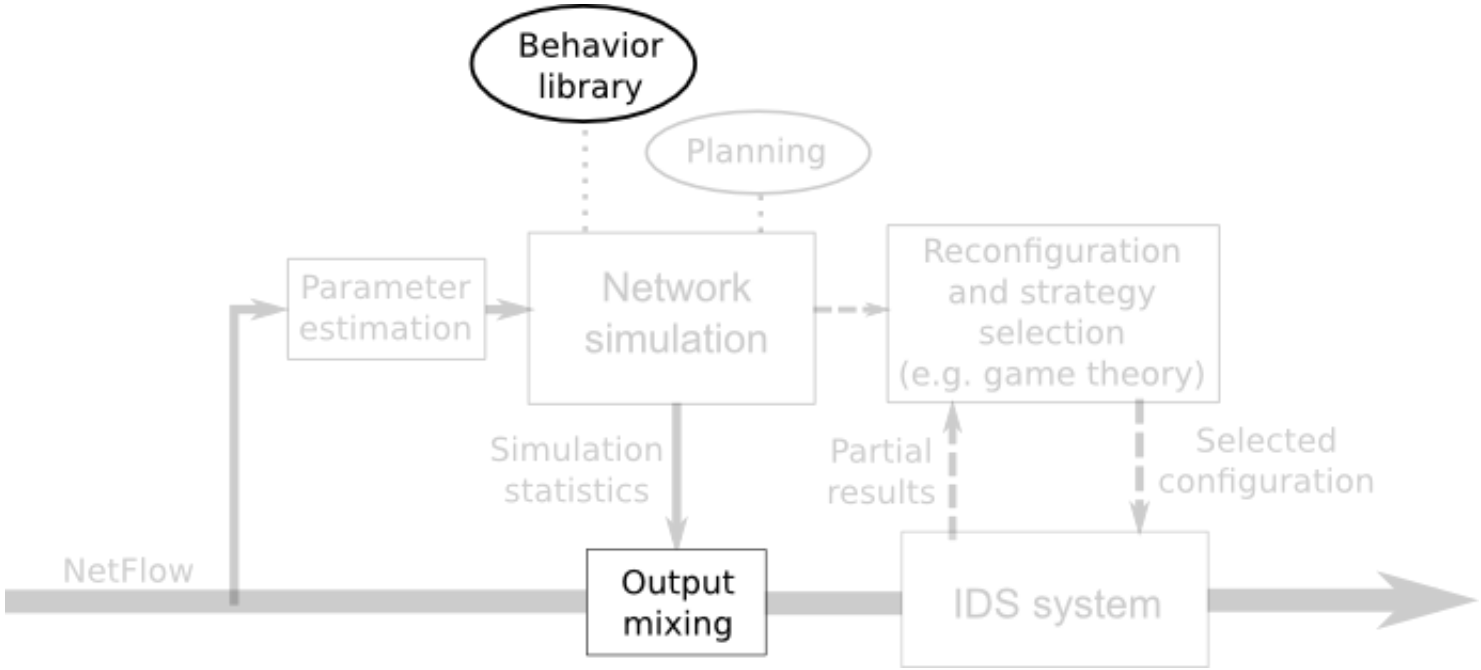
Online Local Adaptation: Research Problems

- Representativeness & Coherence with background traffic
  - Right DNS server IP, representative OS, realistic user profiles of simulated behavior
- Second-order simulation effects (side effects)
- Timeliness

Adaptive Network Simulation

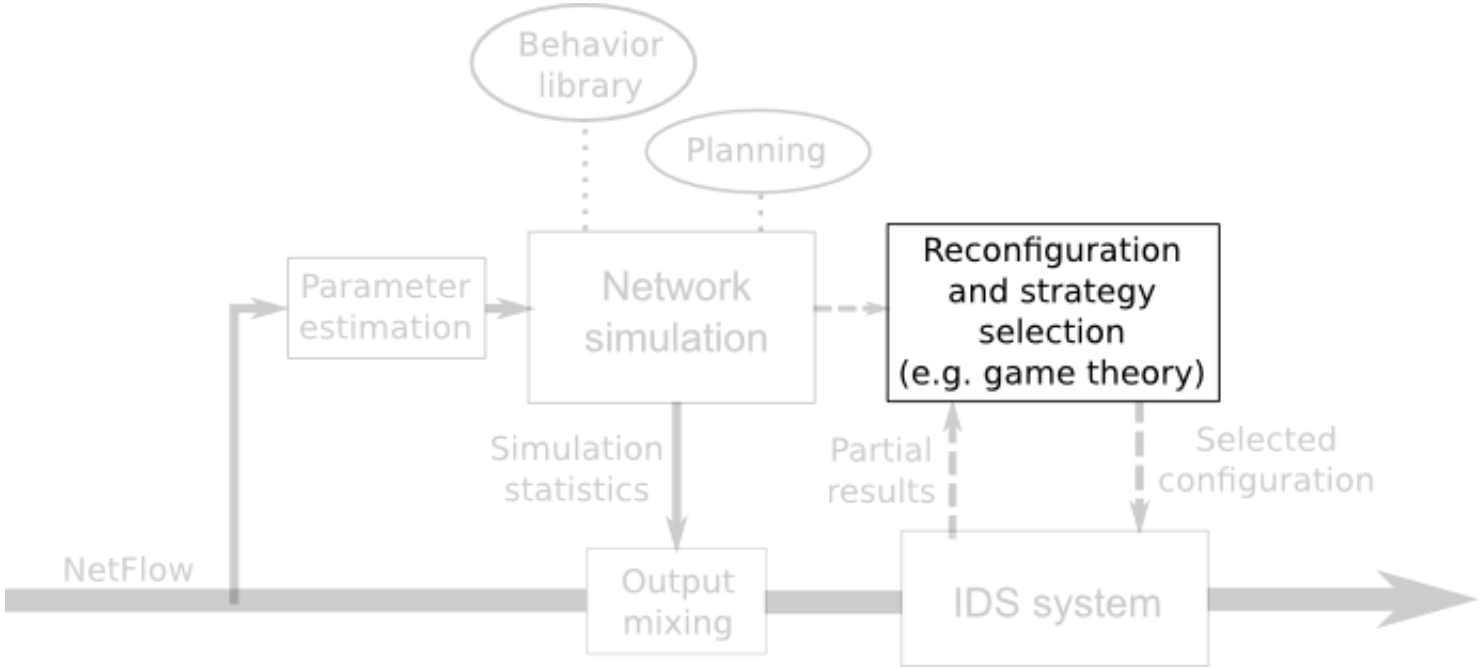


Adaptive Network Simulation



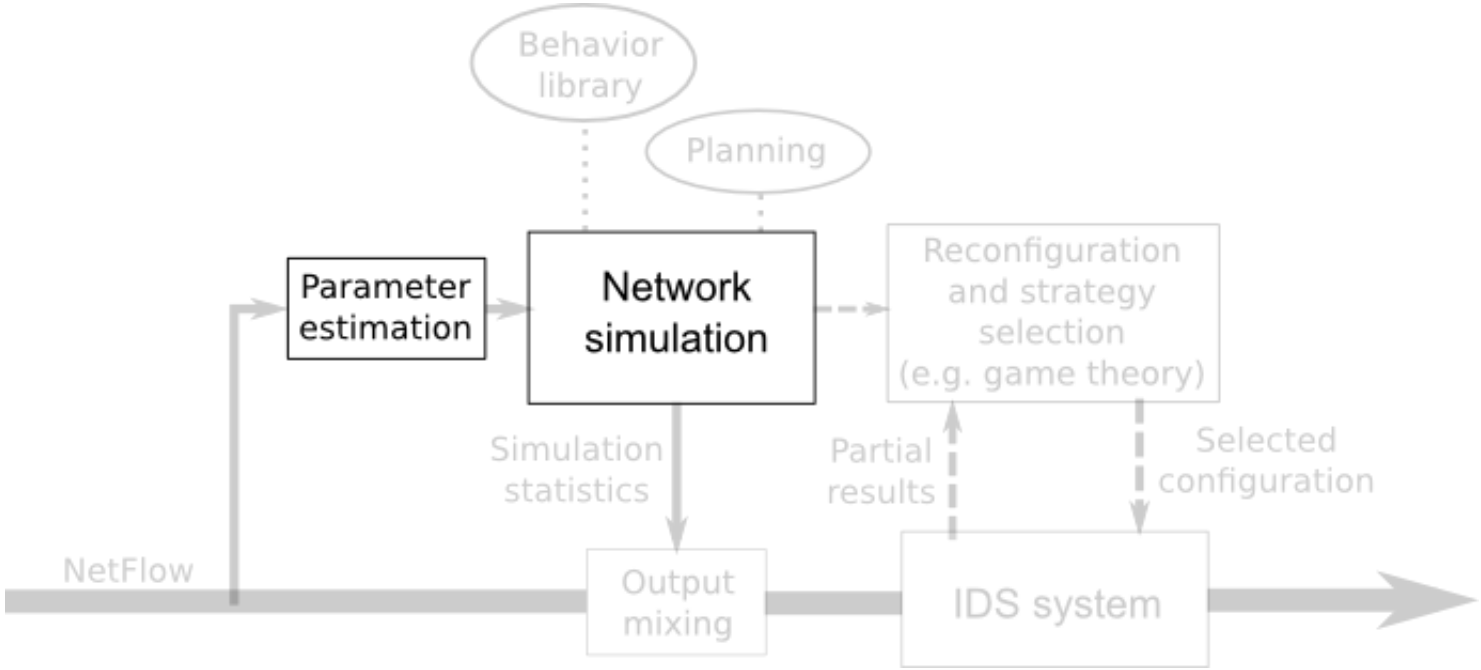
- M. Rehak, M. Pechoucek, M. Grill, J. Stiborek, K. Bartos, and P. Celeda, “Adaptive multiagent system for network traffic monitoring,” Intelligent Systems, IEEE, vol. 24, 2009, pp. 16–25.

Adaptive Network Simulation



- J. Stiborek, M. Grill, M. Rehak, K. Bartos, and J. Jusko, “Game Theoretical Adaptation Model for Intrusion Detection System,” Advances on Practical Applications of Agents and Multi-Agent Systems 2012, pp. 201–210.

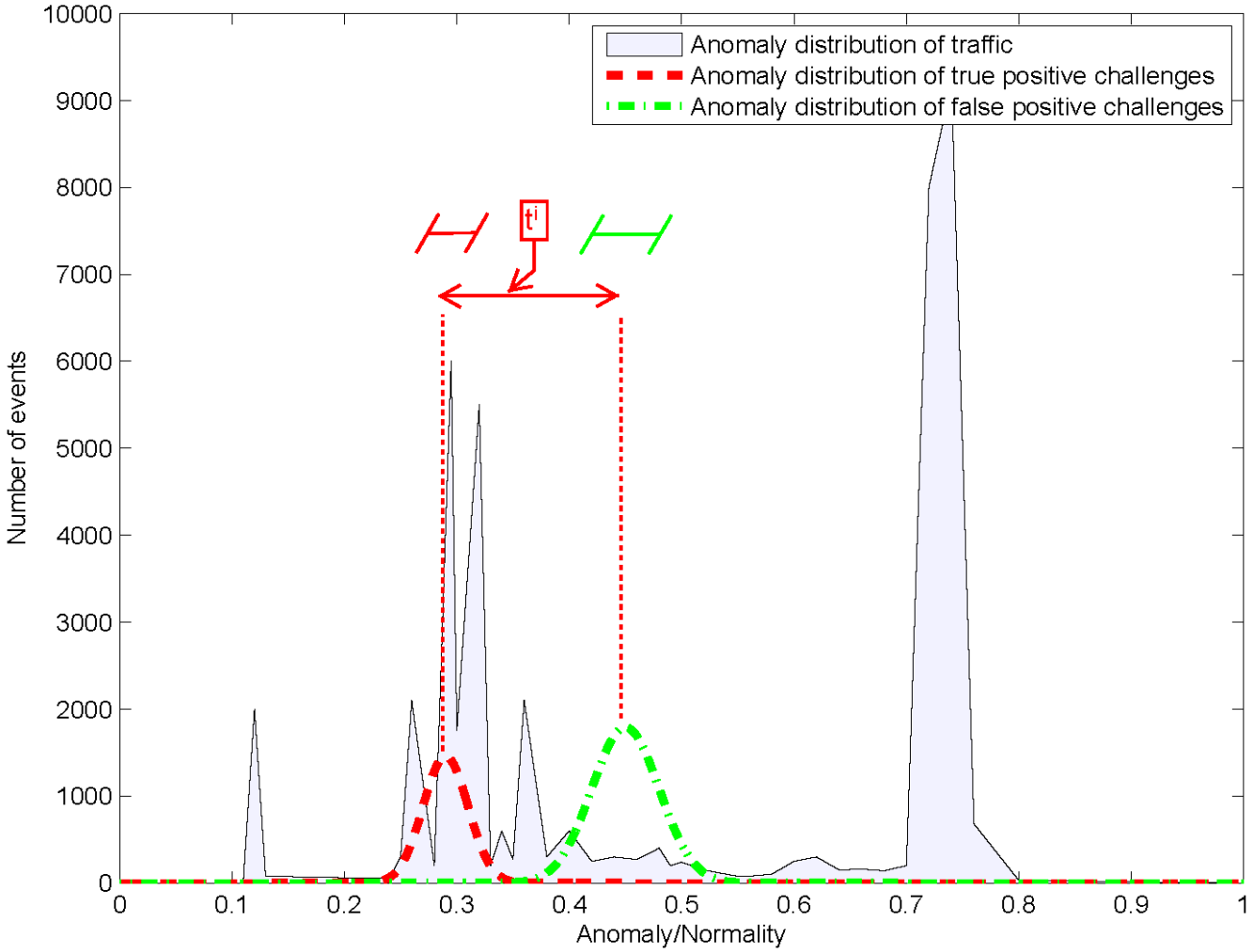
Adaptive Network Simulation



Adaptive Network Simulation

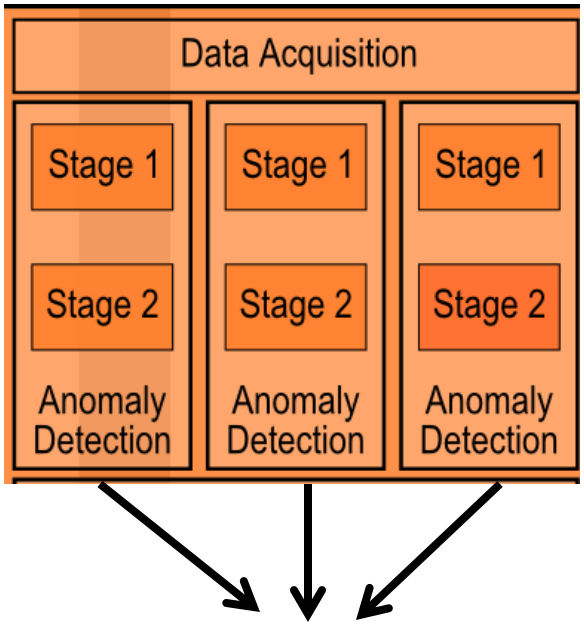
- Estimate joint probability distribution
  - $p(\mathrm{dPort}, \mathrm{dIP}, \mathrm{sPort}, \#\mathrm{bytes}\ (\mathrm{req./resp.}), \#\mathrm{packets}(\mathrm{req./resp.}), \mathrm{thinkTime})$
  - Estimated probability distribution has to follow the correct timeliness
- Generate new samples – Metropolitan Hastings algorithm

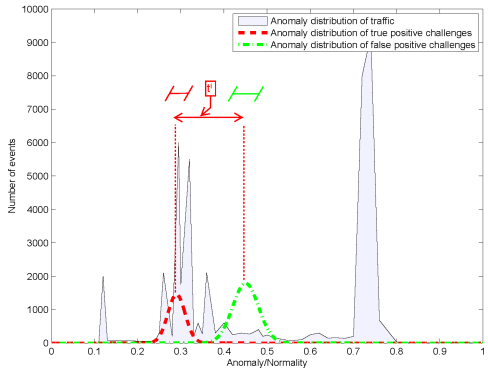
Network Simulation Control



agent technology center

Key Problem n. 2

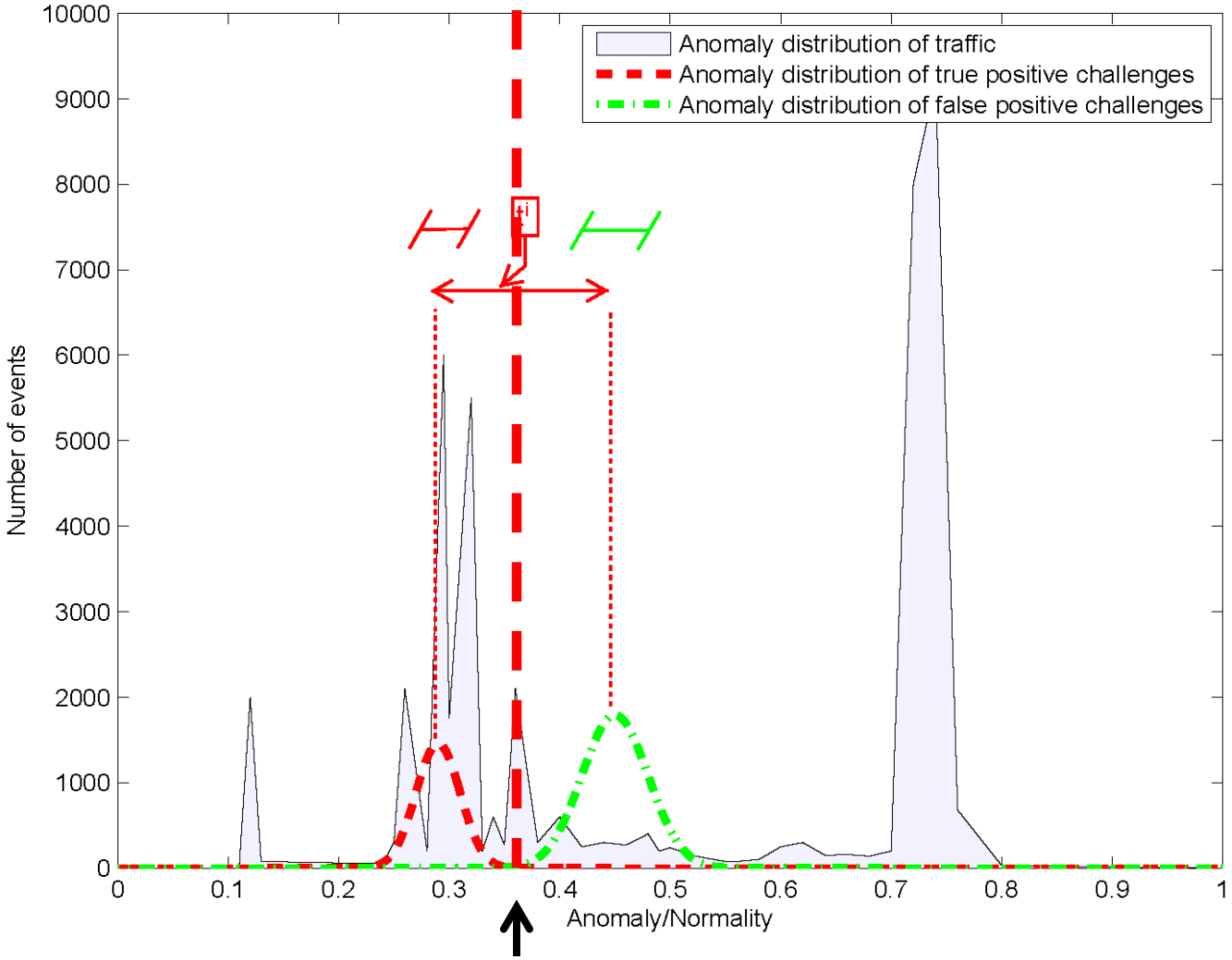




- **If we have aggregated values, where is the threshold separating anomalous and normal traffic?**
- **Solution: Network simulation**

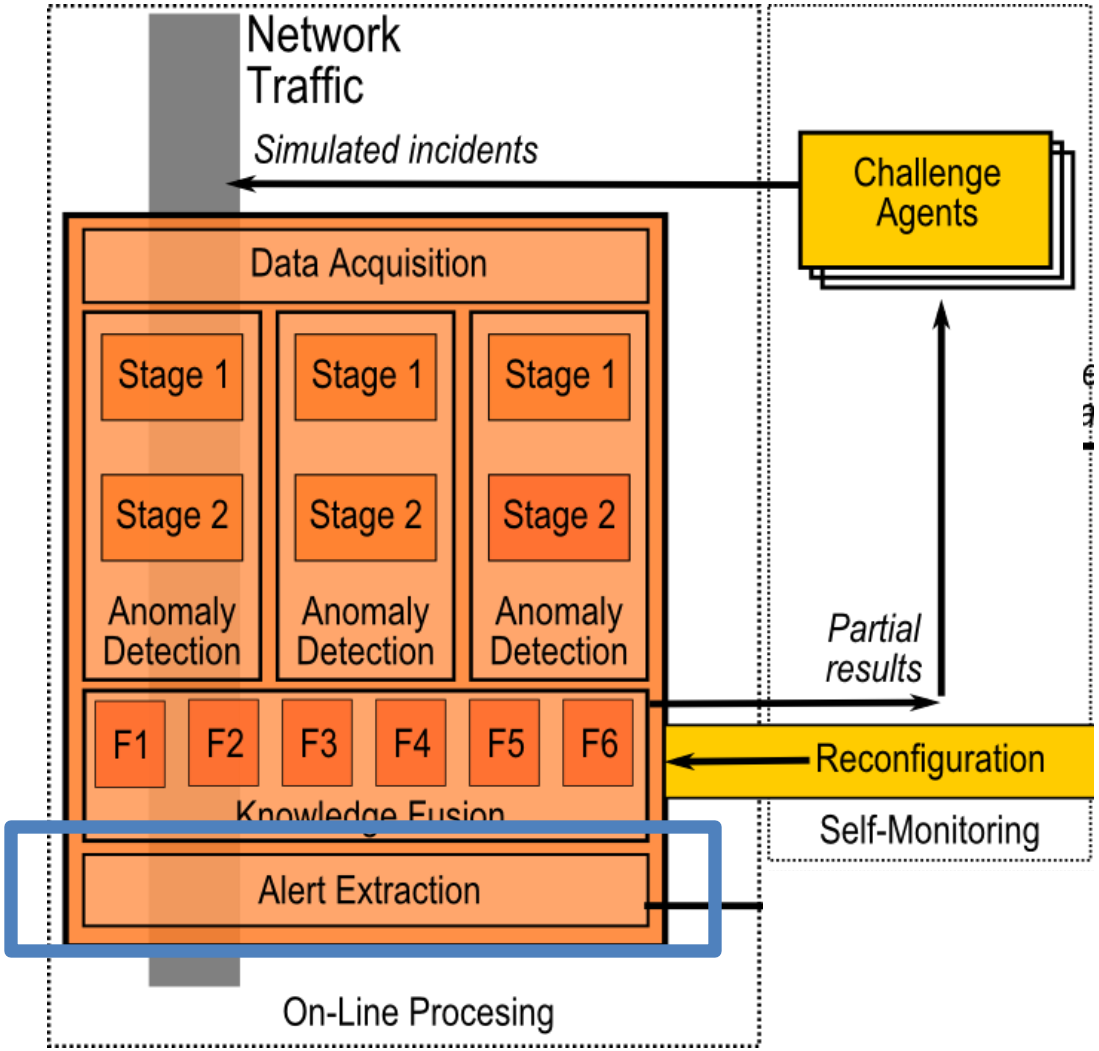
agent technology center

Network Simulation Control



agent technology center

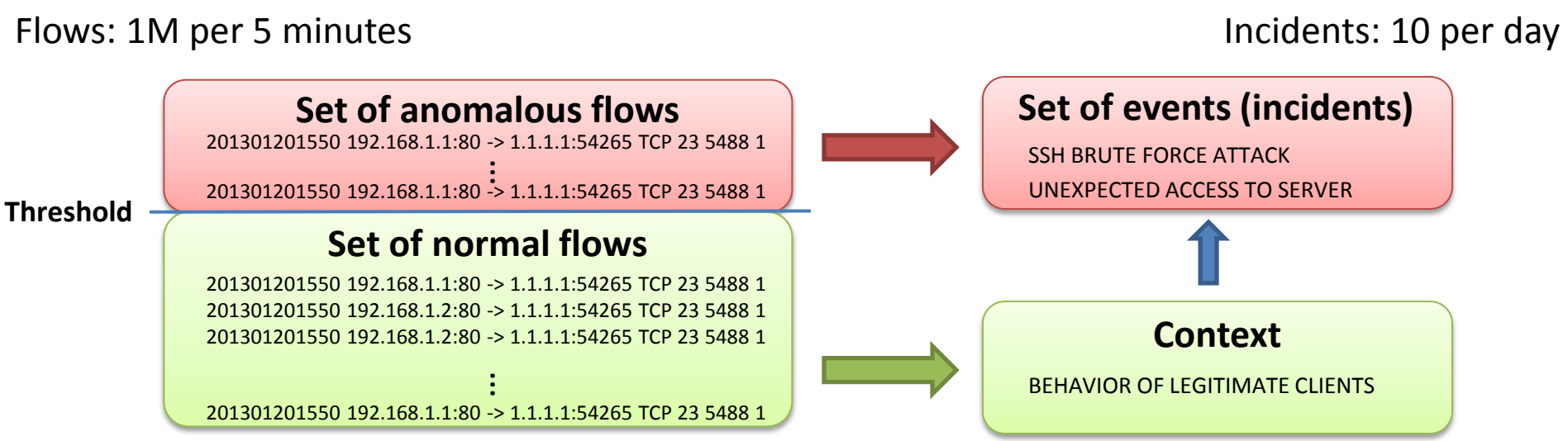
Events Processing Layer



agent technology center

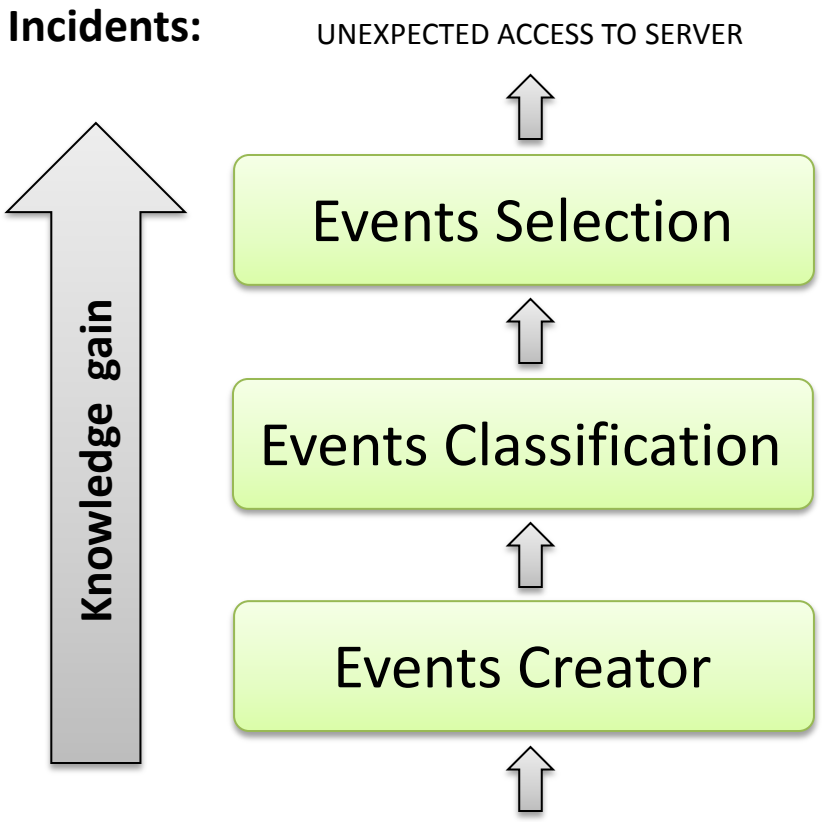
Events Processing Layer

- **Goal: Network security assessments**
  - Increase the level of abstraction
  - Create a list of security incidents
- **Process -** From raw NetFlow data to high-level incidents



agent technology center

Events Processing Layer - Architecture



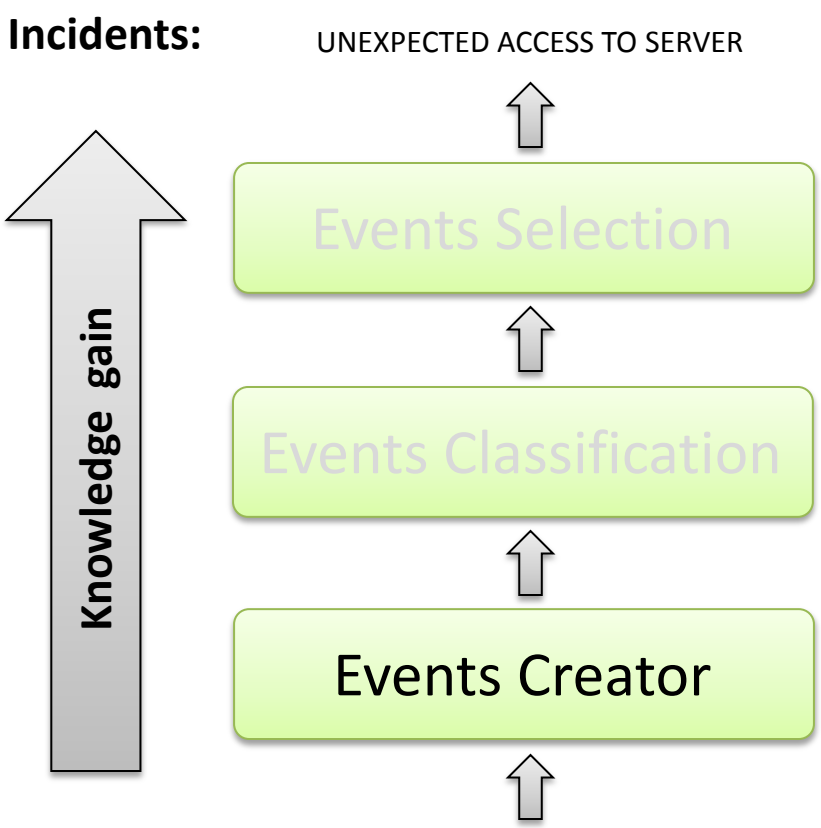
Select and prioritize

Assign label to each event

Group similar flows into cluster => event

agent technology center

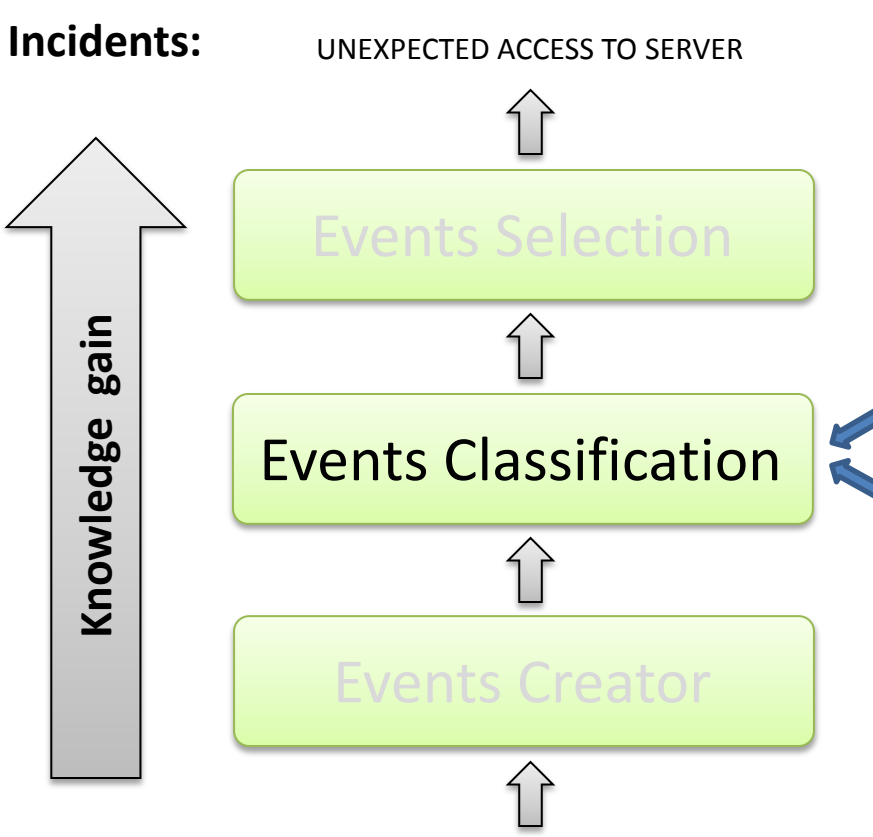
Events Creator



- **Hierarchical clustering**
  - Each cluster represents one type of network behavior
  - Features used in metric:
  - srcIP, srcPrt, dstIP, dstPrt, protocol, bytes
- **Properties**
  - Elementary events
  - High number
  - Small granularity

agent technology center

Events Classification

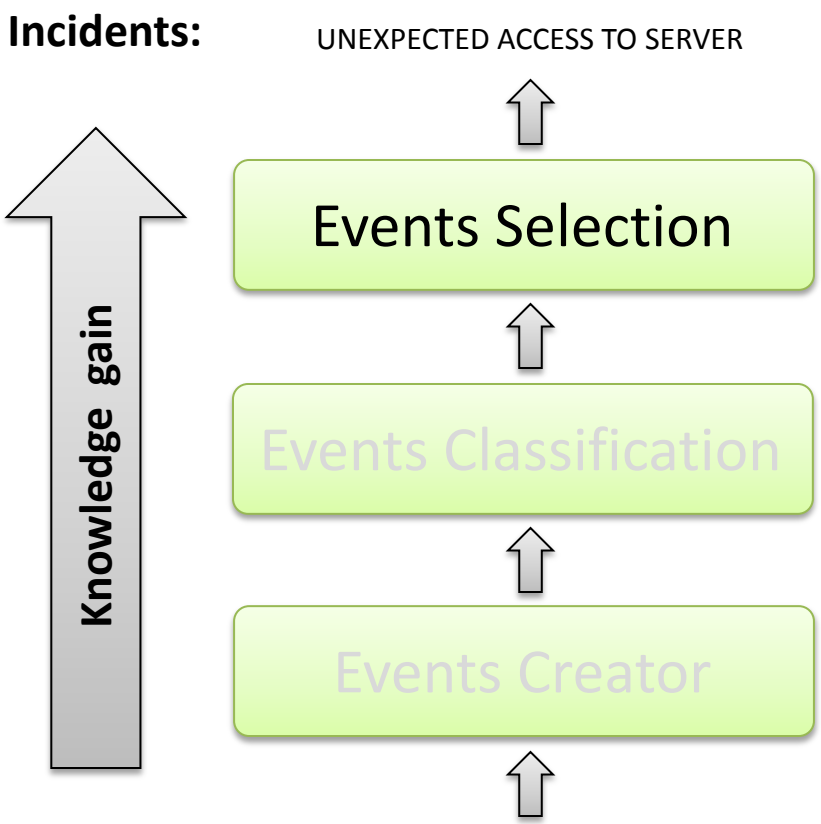


- **Classification**
  - 40 classification agents
  - Agent represents one type of network behavior
  - Form of CNP auction

- **Classification Agents**
  - Stateless – based on expert knowledge (scan)
  - Stateful – uses models
  - P2P, persistence

agent technology center

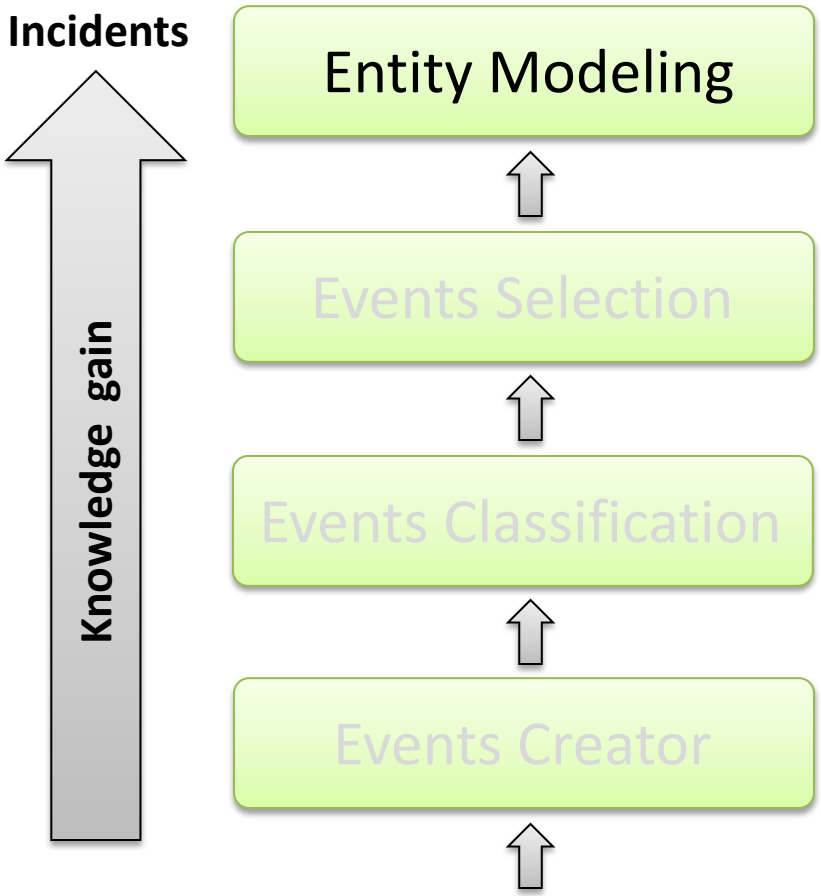
Events Selection



- **Hierarch. Clustering (2nd)**
  - Uses classification and other features
- **Selection & Prioritization**
  - Selects most important events based on their
    - Severity level
    - Degree of anomaly
    - Size (# flows, # bytes)
  - Less important events serve as context

agent technology center

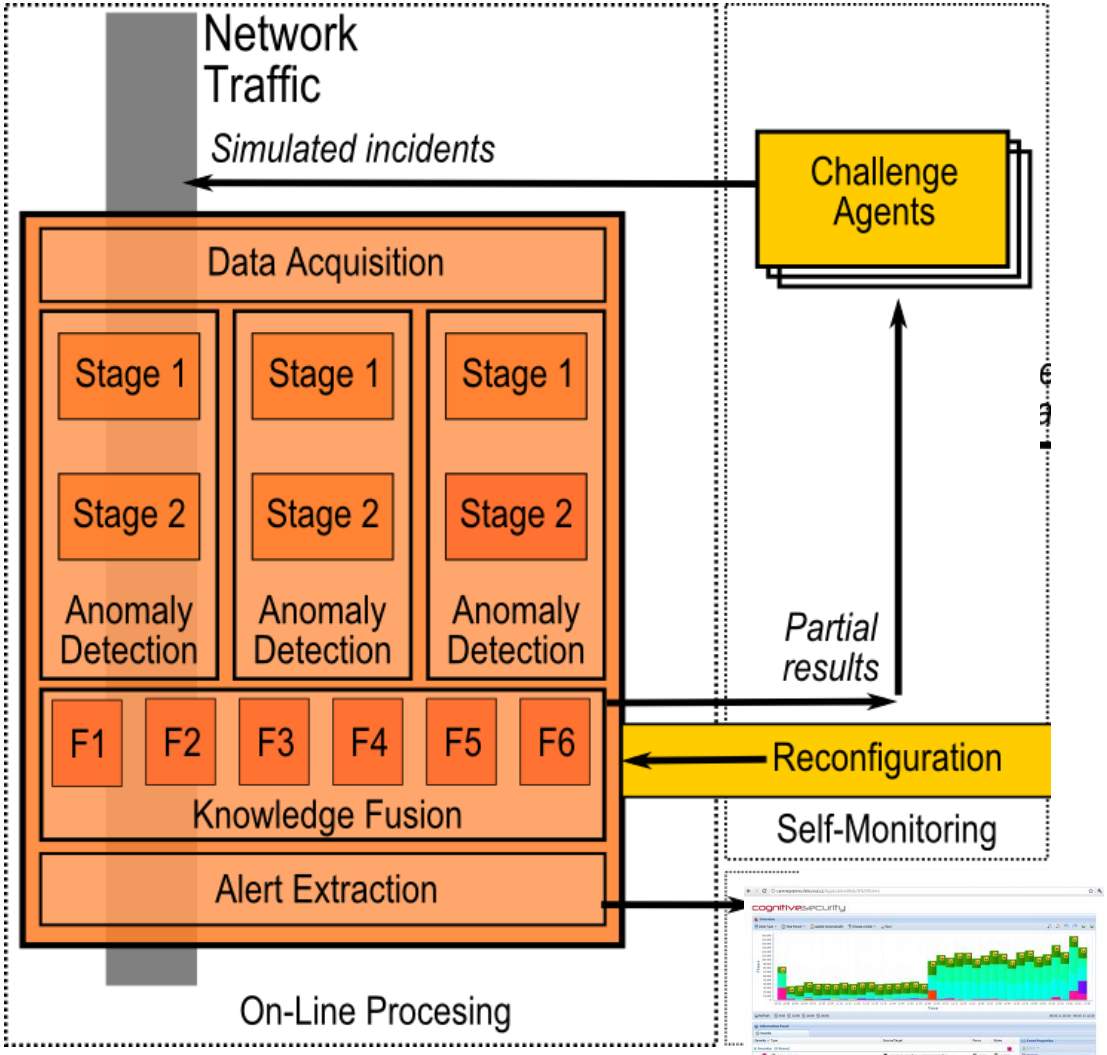
New Component: Entity modeling



- **Entity modeling**
  - Increase the level of abstraction of generated incidents
  - Model the behavior of individual users in time
  - Aggregation and correlation of events
  - Apply supervised learning to known threats to increase precision

agent technology center

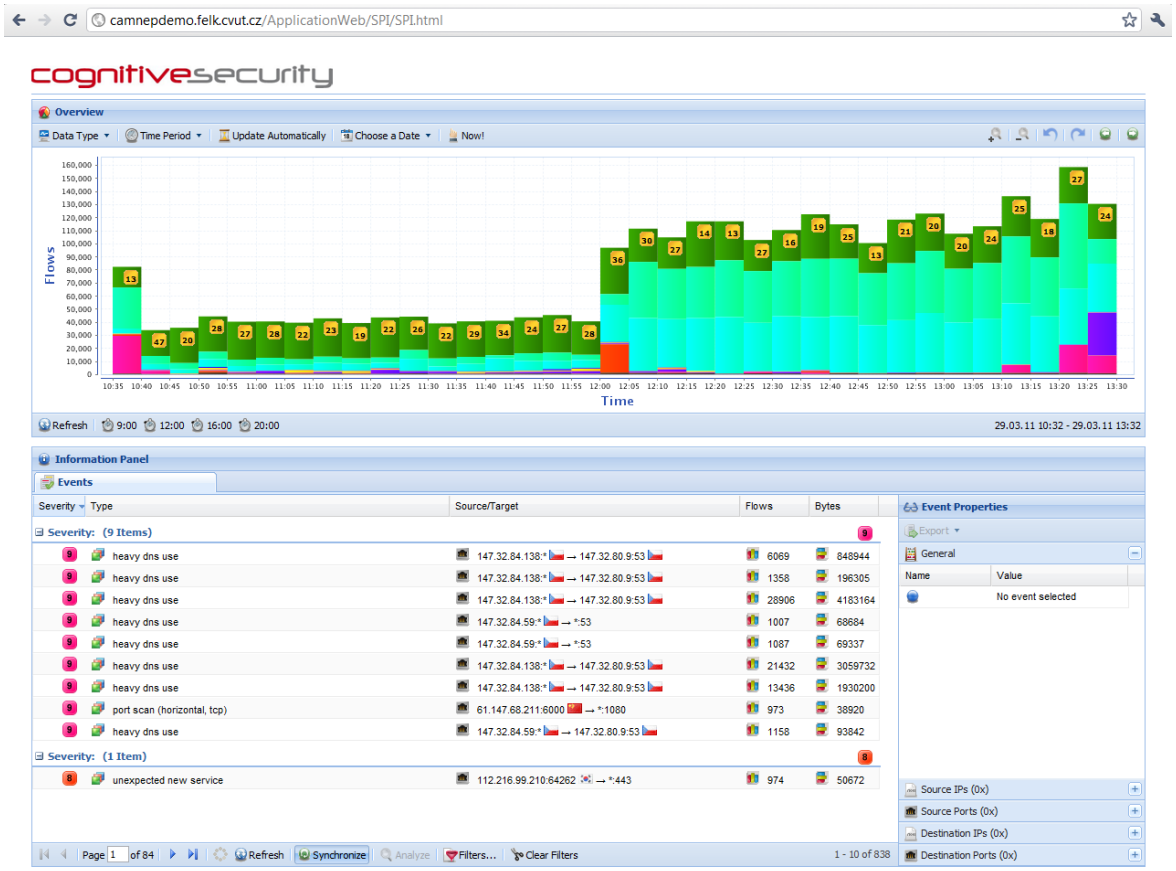
Architecture – Conclusion



- **Data Acquisition and Preprocessing Layer**
  → flows, statistics
- **Detection Layer**
  → trustfulness of flows
- **Self-Monitoring Layer**
  → threshold position
- **Alert Extraction Layer**
  → classified events
- **Analyst Interface**
  → decisions

agent technology center

Demo



agent technology center