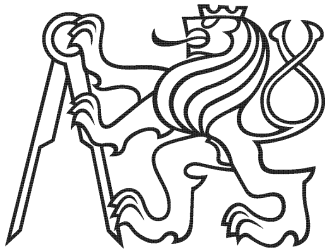


Runtime Monitoring and Dynamic Reconfiguration for Intrusion Detection Systems

Martin Rehak[1,2], Eugen Staab[3], Volker Fusenig[3],
Michal Pechoucek[1,2], Martin Grill[1], Jan Stiborek[1], and Karel Bartos[1]
(1) Czech Technical University in Prague
(2) Cognitive Security
(3) University of Luxembourg

Supported by U.S. ARMY ITC-A/RDECOM – CERDEC project W911NF-08-1-0250

(Research) Questions

What is our IDS/NBA good for ?

Does it work right now ?

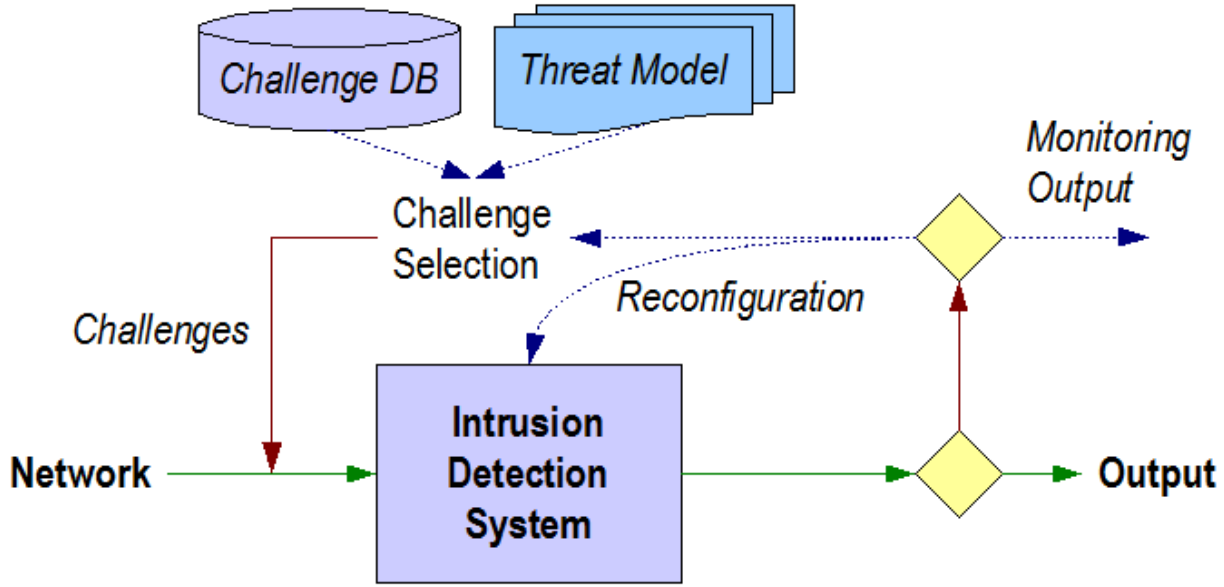
How sensitive it is ?

Can it detect X ?

Our Answer

Use of **trust modeling** techniques combined with **challenge insertion** for a dynamic reconfiguration of an anomaly-based **network intrusion detection** system

Challenge-based Monitoring



- Unlabeled background input data
- Insertion of small set of challenges
  - Legitimate vs Malicious

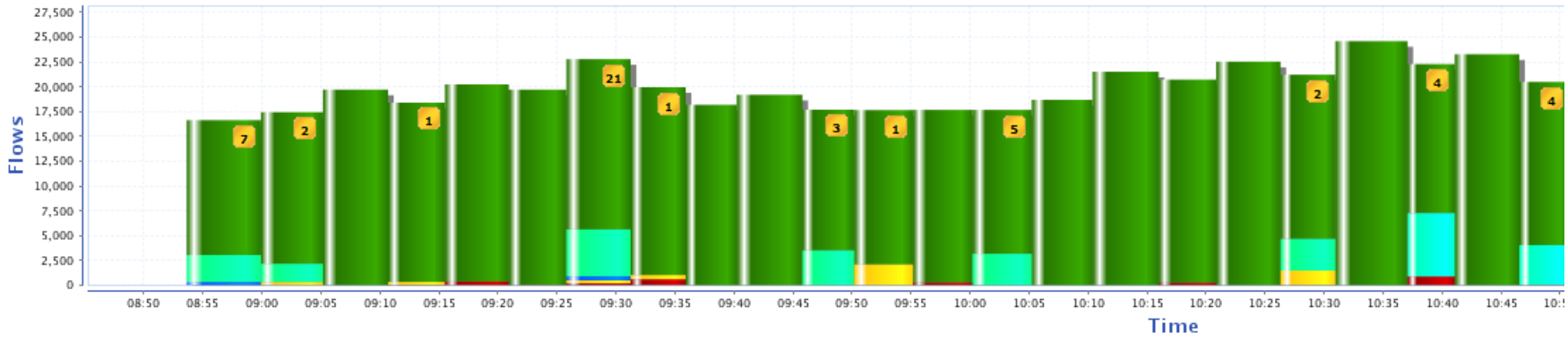
(1) Response evaluation

(2) What challenges ?

(3) How many ?

Overview

Data Type | Time Period | Update Automatically | Choose a Date | Now!



Refresh | 9:00 | 12:00 | 16:00 | 20:00

Events

Start Time	Severity	Type	SimilarType	Flows	Bytes	Protocols	Source/Target
Protocols: TCP (68 Items)							
Sep 23, 11:25:28	8	port_scan_horizontal		956	38240	TCP	123.215.231.250:6000 -> *:1433
Sep 23, 11:23:33	5	Default Incident Type Agent		959	38360	TCP	58.254.213.70:6000 -> *:53
Sep 23, 11:11:14	5	Default Incident Type Agent		4877	163668810	TCP	147.32.80.13:* -> *:*
Sep 23, 08:59:44	5	Default Incident Type Agent		286	195154	TCP	79.112.106.28:* -> 147.32.84.2:*
Sep 23, 09:30:59	5	Default Incident Type Agent		449	758999	TCP	147.32.85.34:* -> *:*
Sep 23, 10:28:24	5	Default Incident Type Agent		1468	58720	TCP	60.173.26.47:6000 -> *:*
Sep 23, 09:49:53	5	Default Incident Type Agent		2079	83751529	TCP	147.32.80.13:* -> *:*
Sep 23, 09:10:37	5	Default Incident Type Agent		297	199069	TCP	94.21.179.58:* -> 147.32.84.2:*
Sep 23, 09:25:49	5	Default Incident Type Agent		301	1369955	TCP	147.32.84.2:* -> *:*
Sep 23, 08:53:51	3	ftp_requests		122	42023	TCP	115.95.237.83:* -> 147.32.85.34:*
Sep 23, 09:25:48	3	ftp_requests	smtp_requests	121	20817	TCP	147.32.84.194:* -> *:*
Sep 23, 11:21:34	3	sql_requests_server	port_scan_horizontal	67	2680	TCP	*:80 -> *:*
Sep 23, 08:53:53	3	ftp_responses		176	714739	TCP	147.32.85.34:* -> *:*

Refresh | Synchronize

Anomaly Detection vs. Signatures

Signature matching

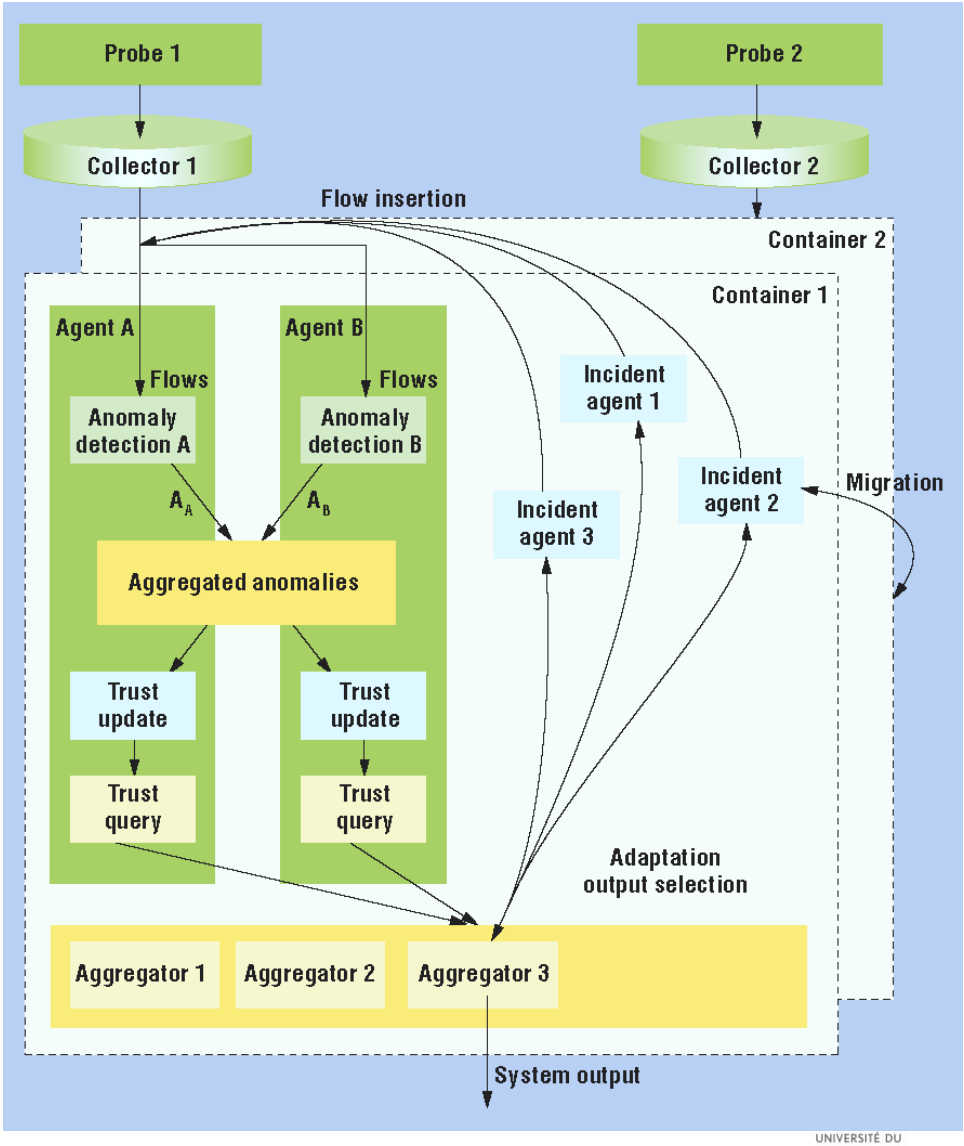
- Historically validated
- Widely deployed
- Verifiable & Stable
- Number of patterns
- Scaling
- Management
- New threats detection

Anomaly detection

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

CAMNEP: Detection Layer

- **Flows** to **categories**
- Multiple AD methods
- Multiple trust models
- Multiple aggregation methods
- Dynamic
- Several layers of learning

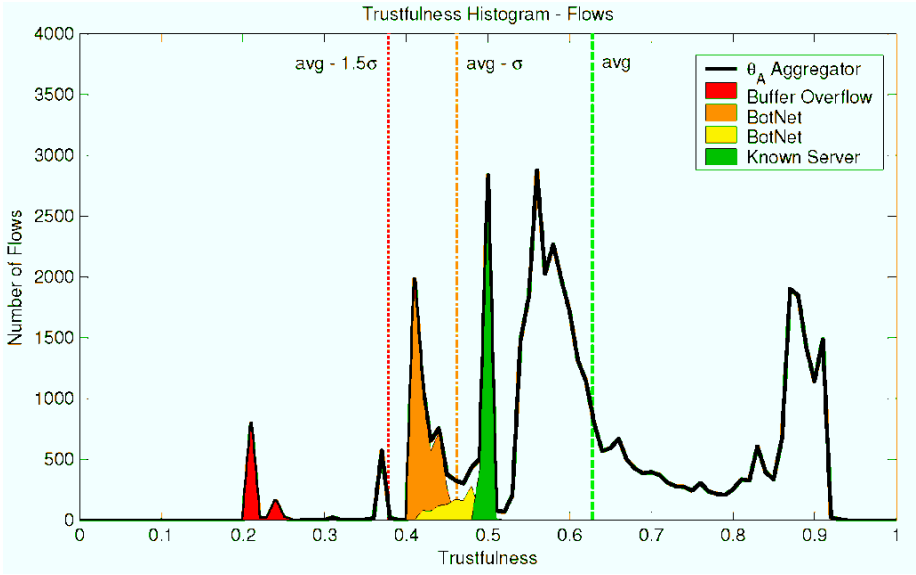


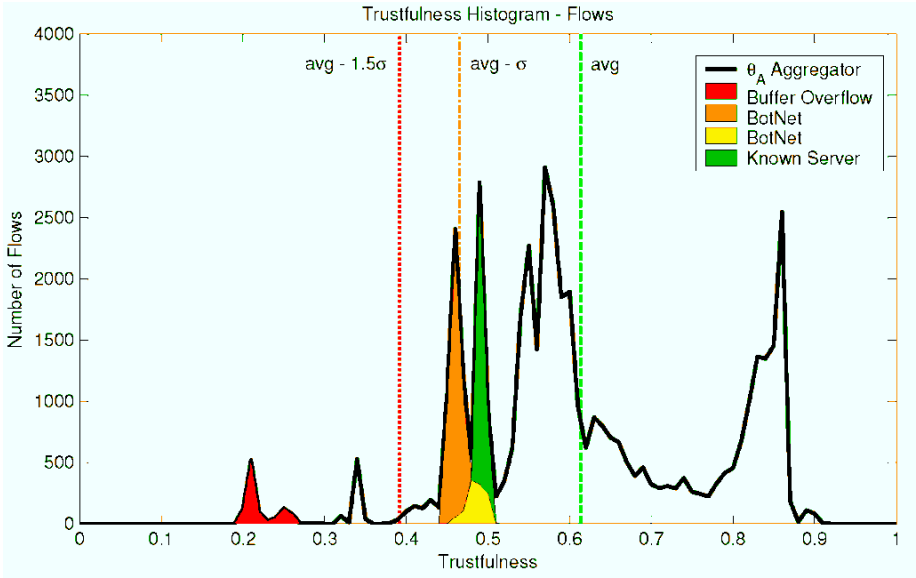
agent technology center

UNIVERSITÉ DU LUXEMBOURG

Dynamic classifier selection

- Unsupervised
- Dynamic
  - Background traffic
  - Model performance
  - Attacks
- Strategic behavior
  - Evasion
  - Attacks on AD/learning





agent technology center

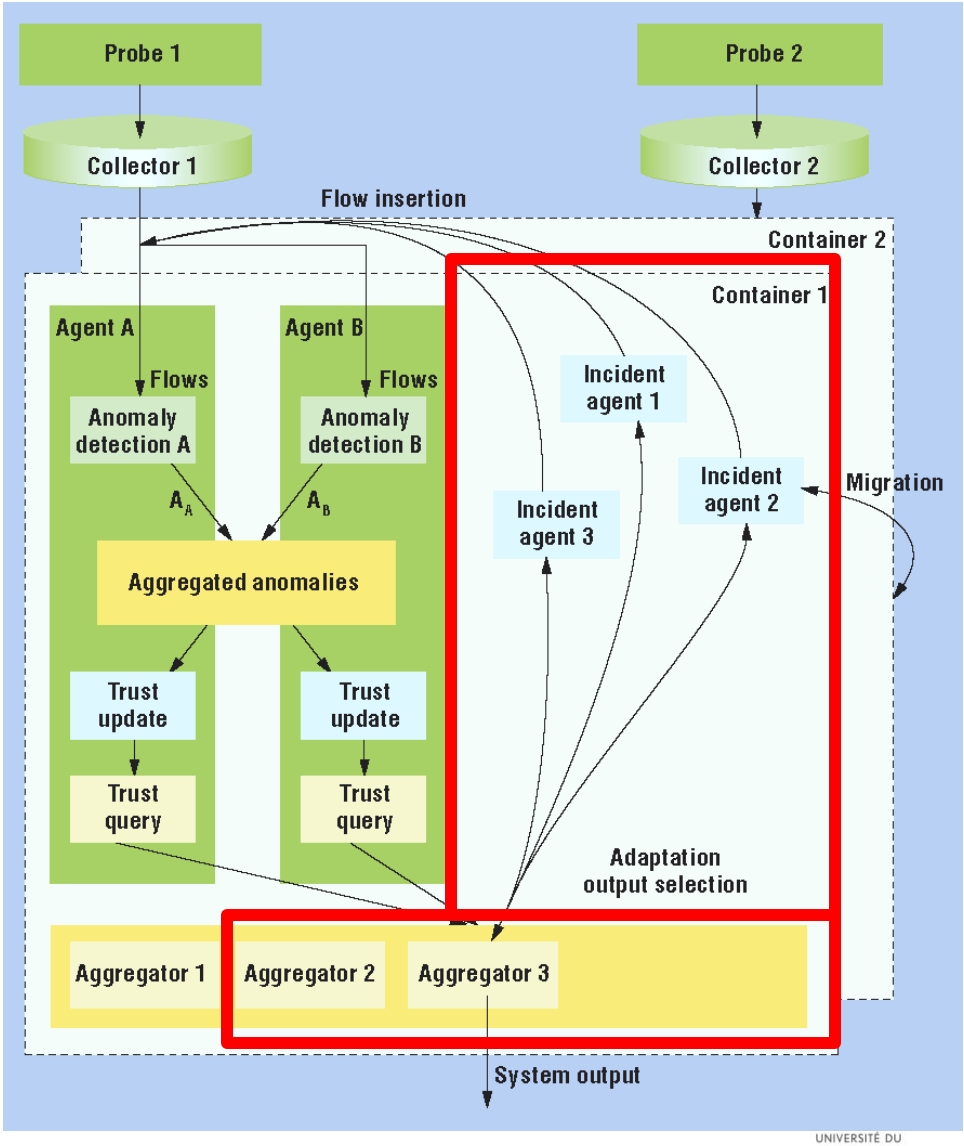
Why bother ?: False/True Positives

Individual AD methods 300:2

Averaged anomalies 58:2

Averaged trust 15:2

Adaptive average 5:2

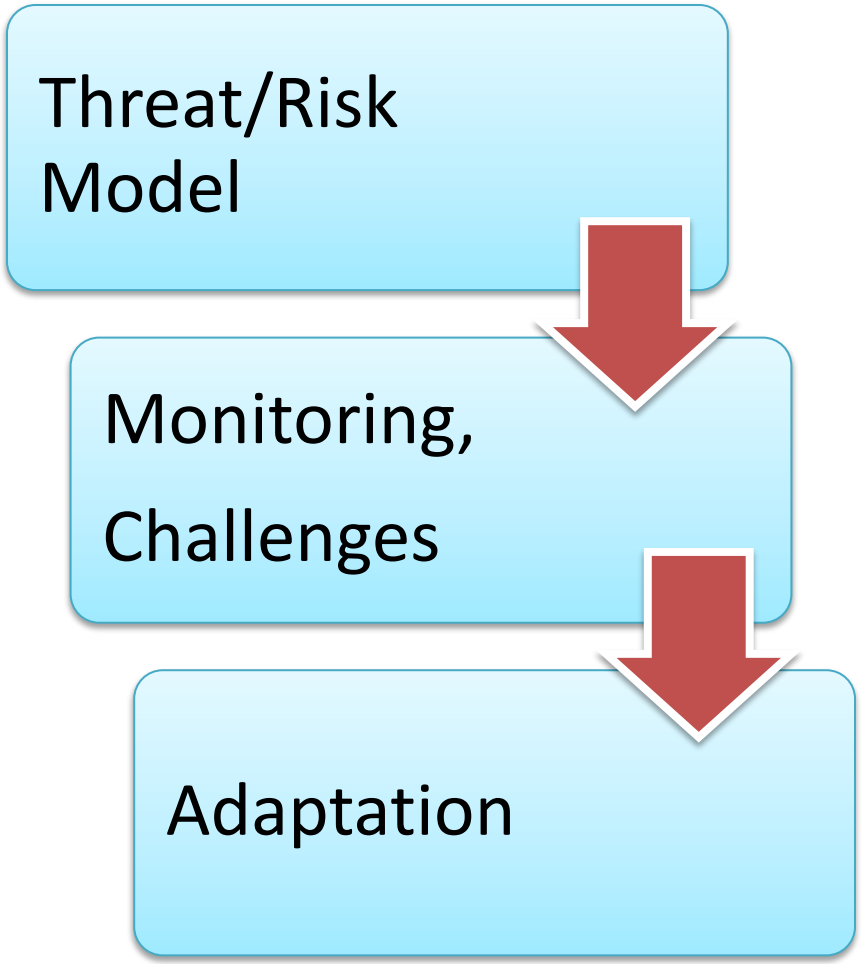


agent technology center

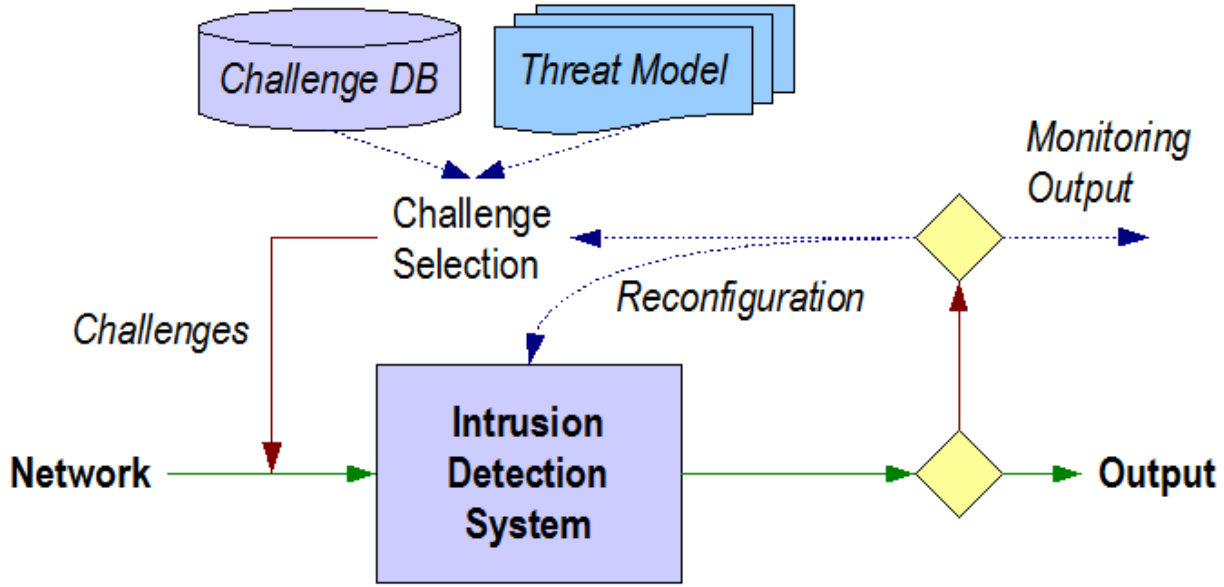
UNIVERSITÉ DU LUXEMBOURG

Adaptation Principles

- Self-Awareness:
  - Self-monitoring
  - Self-evaluation
  - Goal representation
- Self-Optimization:
  - (Aggregation generation)
  - Aggregation function selection



Monitoring: Challenge Insertion



- Unlabeled background input data
- Insertion of small set of challenges
  - Legitimate vs Malicious

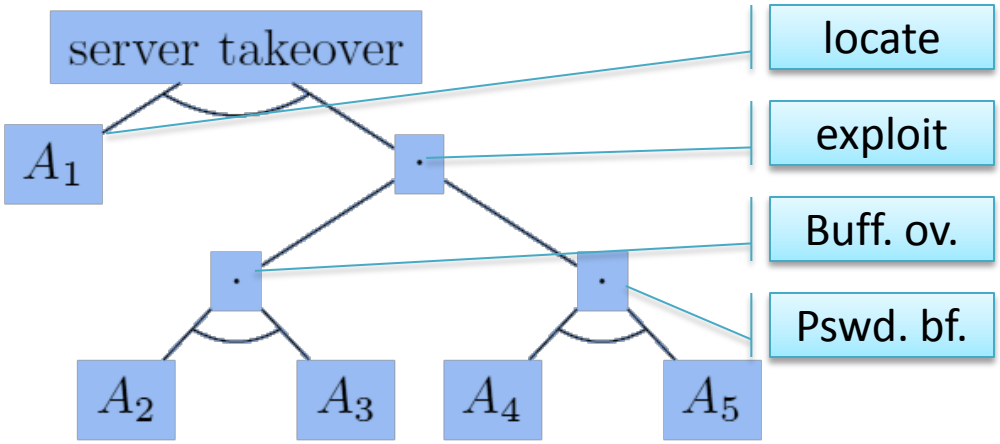
(1) Response evaluation

(2) What challenges ?

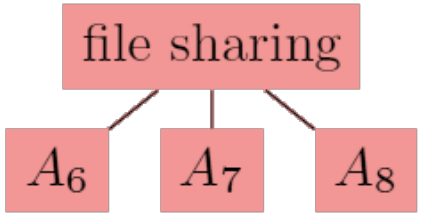
(3) How many ?

Attack Trees - (Simplified) Examples

Server take-over



File sharing



A_1	horizontal scan
A_2	fingerprinting
A_3	buffer overflow
A_4	SSH brute force request
A_5	SSH brute force response

A_6	download
A_7	upload
A_8	directory node

Decision-Theoretic Threat Modeling

- Threat modeled as:
  - attack tree (T)
  - **loss value** (D)
- Loss values propagation to leaf nodes, i.e. attack actions (A_i)
- Loss value aggregated over threats for **attack classes** (AC)

$$F(T) = \{\{A_1\}, \{A_2, A_3\}, \{A_2, A_4\}\}$$

$$P(A_i, T_j) := \frac{1}{|F(T_j)|} \sum_{\substack{C_k \in F(T_j), \\ \text{with } A_i \in C_k}} \frac{1}{|C_k|}$$

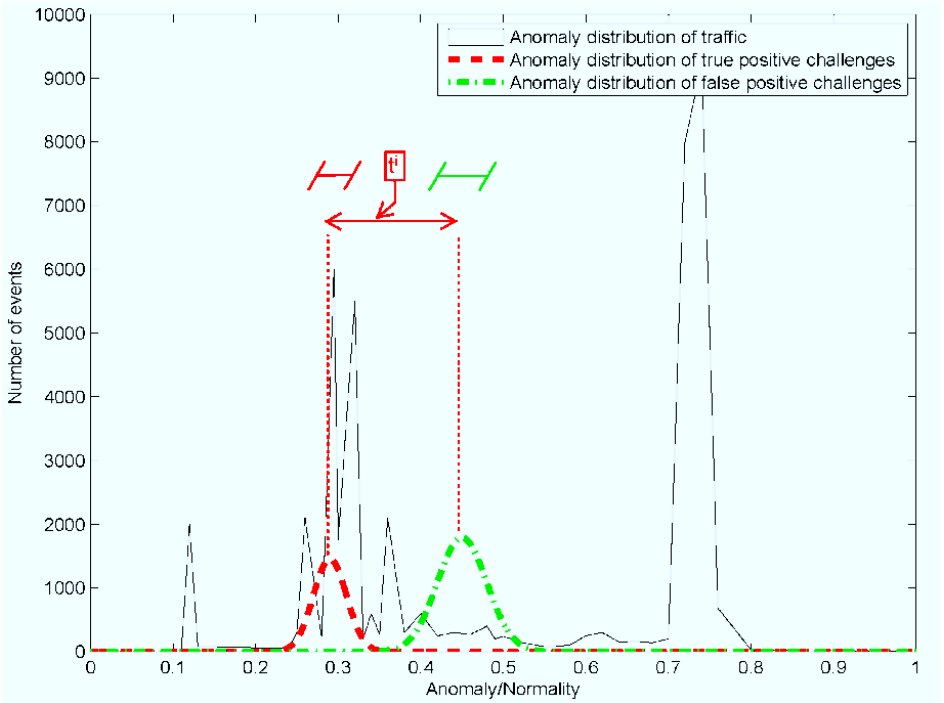
$$P(A_i) := \frac{1}{\sum_{T_j \in \mathcal{T}} D(T_j)} \cdot \sum_{T_k \in \mathcal{T}} D(T_k) \cdot P(A_i, T_k)$$

$$P(AC) = \sum_{A_i \in AC} P(A_i)$$

From Challenge Insertion to Trust

- **Trust** in the aggregator agent models its ability to separate the legitimate from malicious behavior under current conditions

$$t_{\alpha}^{i,k} = \frac{\bar{y} - \bar{x}^{k}}{\sigma_y + \sigma_x^k}$$



Trust Modeling – Issues

- Regret/FIRE model individual reputation component used
  - Startup delay considerations
  - Changing network traffic character
  - Number of inserted challenges vs. the number of attack types
  - Relationship between the challenge insertion and trust

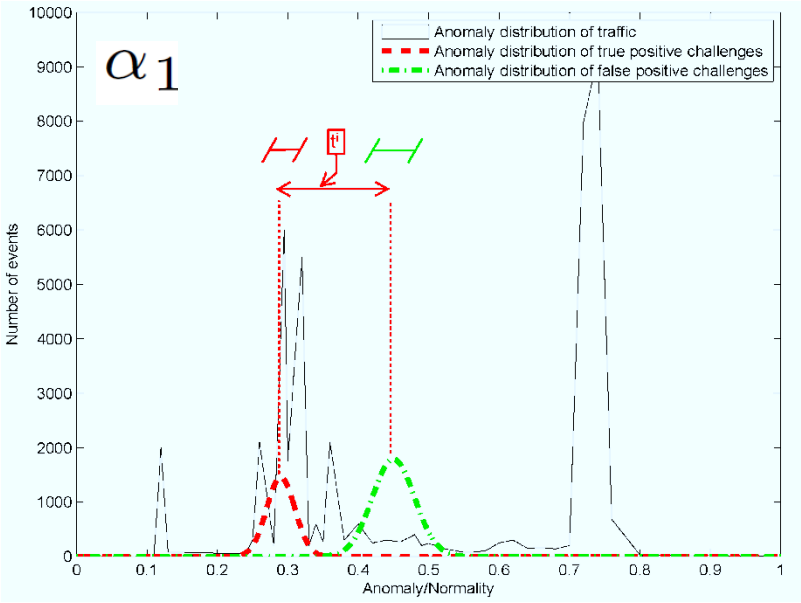
$$t_{\alpha}^{i,k} = \frac{\bar{y} - \bar{x}^{k}}{\sigma_y + \sigma_x^k} \qquad w_i = \frac{1}{W} e^{(j-i)\frac{\ln(0.1)}{4}}$$

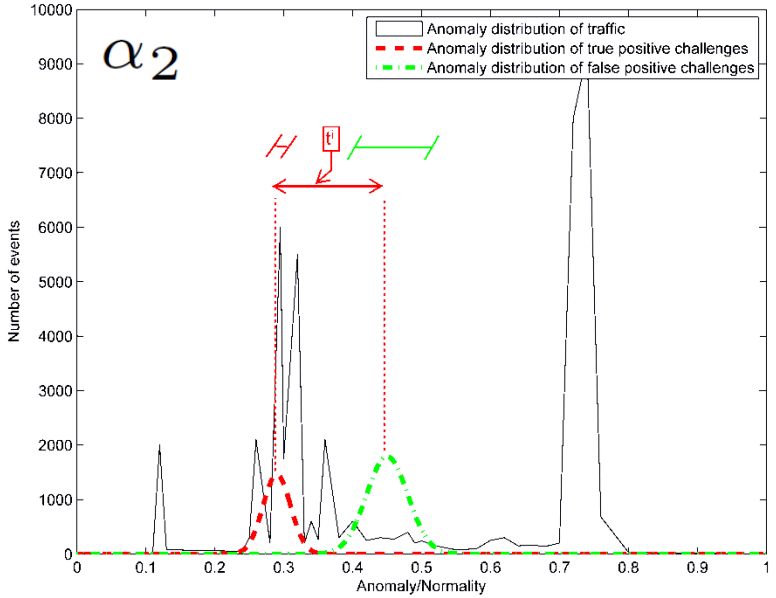
$$T_{\alpha}^{k} = \sum_{i} w_i * t_{\alpha}^{i,k}$$

Challenge Insertion Control

$$T_\alpha^k = \sum_i w_i * t_\alpha^{i,k}$$

$$T_{\alpha_1} \geq T_{\alpha_2}$$





Challenge Insertion Control (2)

$$n = \left(\frac{z^* \sigma_x^k}{m} \right)^2 \Longrightarrow T_{\alpha_1} \geq T_{\alpha_2}$$

$$m' = \frac{(t_{\alpha_1} - t_{\alpha_2})ab}{2(a+b)} = \frac{b(\bar{y}_1 - \bar{x}_1^k) - a(\bar{y}_2 - \bar{x}_2^k)}{2(a+b)}$$

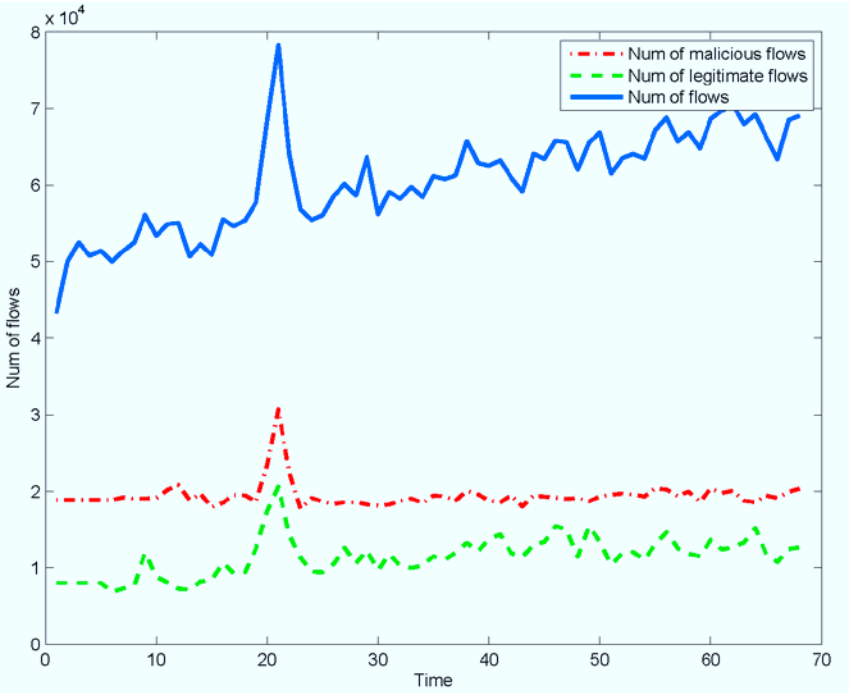
$$t_{\alpha_1} \geq \frac{\bar{y}_1 - \bar{x}_1^k - 2m'}{\underbrace{\sigma_{y_1} + \sigma_{x_1}}_{=:a}^{k}} = \frac{\bar{y}_2 - \bar{x}_2^k + 2m'}{\underbrace{\sigma_{y_2} + \sigma_{x_2}}_{=:b}^{k}} \geq t_{\alpha_2}$$

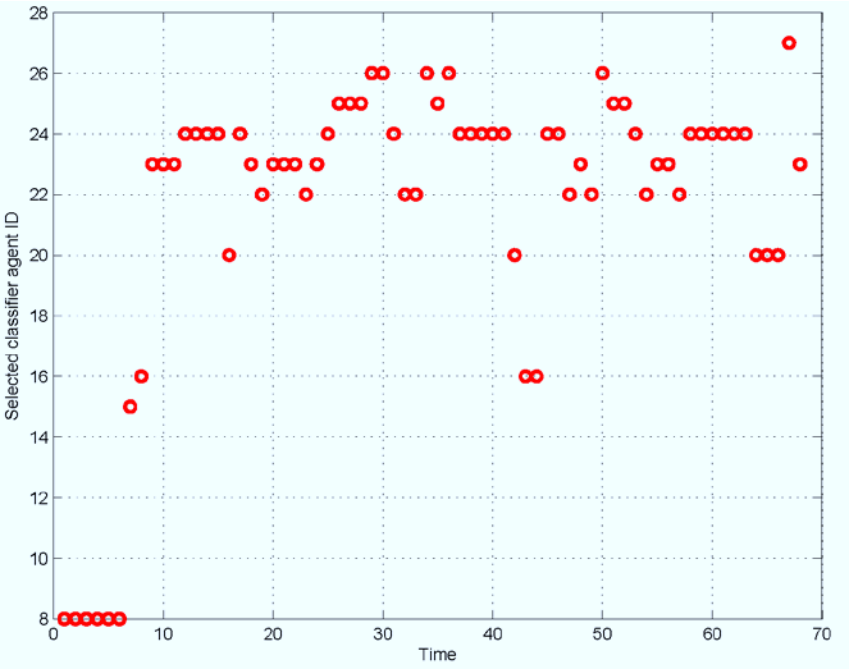
- Trust values used to parameter the challenge insertion
- We prevent random order inversion between the two most trusted agents

Evaluation

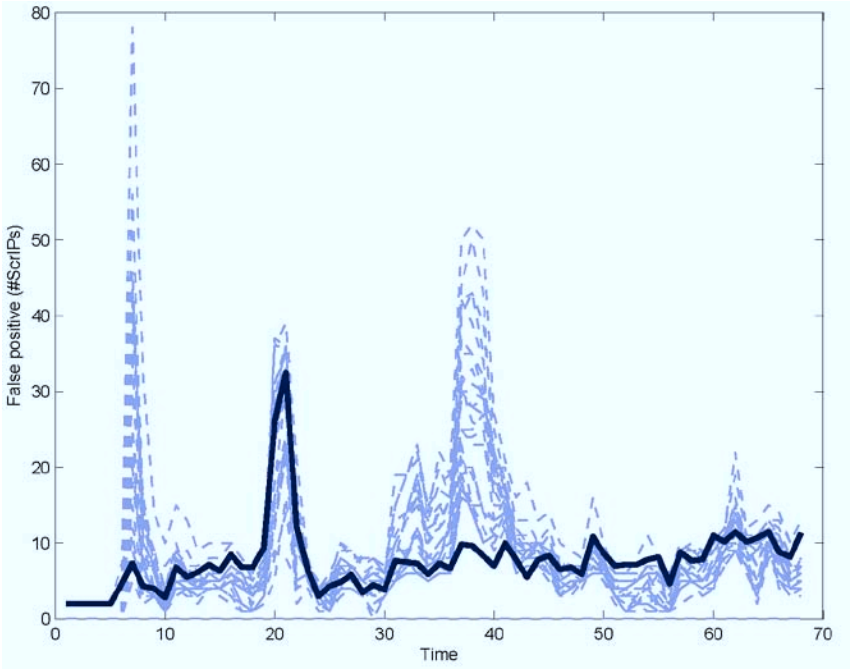
- Real network traffic
  - 1Gb link
  - 200-300 Mb/sec eff.
  - 200 flows/sec
  - 6 hours … 70 datasets
  - 5 minute collection
- Third party attacks
- SSH scans, password brute force, worms/botnets, malware, P2P

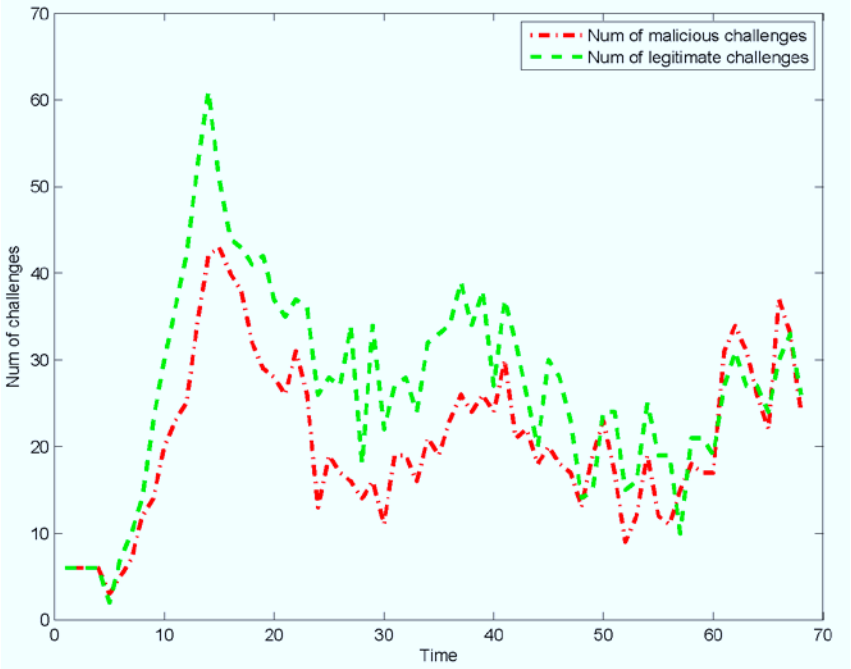
Experimental results





Experimental results





Experimental Results

- False positives reduced (excesses avoided)
- False negatives comparable/reduced

Aggregation	False Negative (sIP)	False Positive (sIP)
Arithmetic average	14.7	12.5
Average aggregation fct.	13.1	24.3
Min FP aggregation fct.	14.5	5.3
Min FN aggregation fct.	9.8	125.2
Best aggregation fct.	13.7	5.7
Adaptive selection	14.0	3.1

- University network, third party attacks only – scans, P2P, password bf,…

Attack-Type Insertion Effects

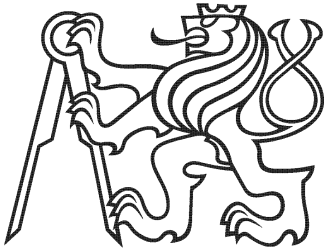
- Observable effects on trustfulness values
- Slow/low volume attacks are still undetectable
- So far inconclusive on the extracted event level
- Natural background traffic, known test attacks

Attack	All challenges	Selected challenges
Horizontal scan	1.1/-0.2	1.4/0.0
Vertical scan	1.2/-0.2	1.4/0.3
Fingerprinting	1.5/1.2	1.9/1.6
SSH pass. brute force	-0.2/0.6	0.17/1.2
Buffer overflow	-0.2/0.1	0.2/0.0

Conclusions

- Advanced AI techniques can:
  - Automatically reduce and maintain the **error rate**
  - Monitor system **performance**
  - **Optimize** system performance by:
    - Aggregation function selection
    - Challenge insertion process management
- Current/Future work
  - behavior generation (promising)
  - reduction of evasion/strategic behavior
  - opponent models

Questions ?



cognitivesecurity

UNIVERSITÉ DU
LUXEMBOURG

rehak@cognitivesecurity.cz