Intrusion Prevention through Optimal Stopping

Digital Futures Machine Learning Day, 17 January 2022

Kim Hammar (kimham@kth.se) Rolf Stadler (stadler@kth.se)

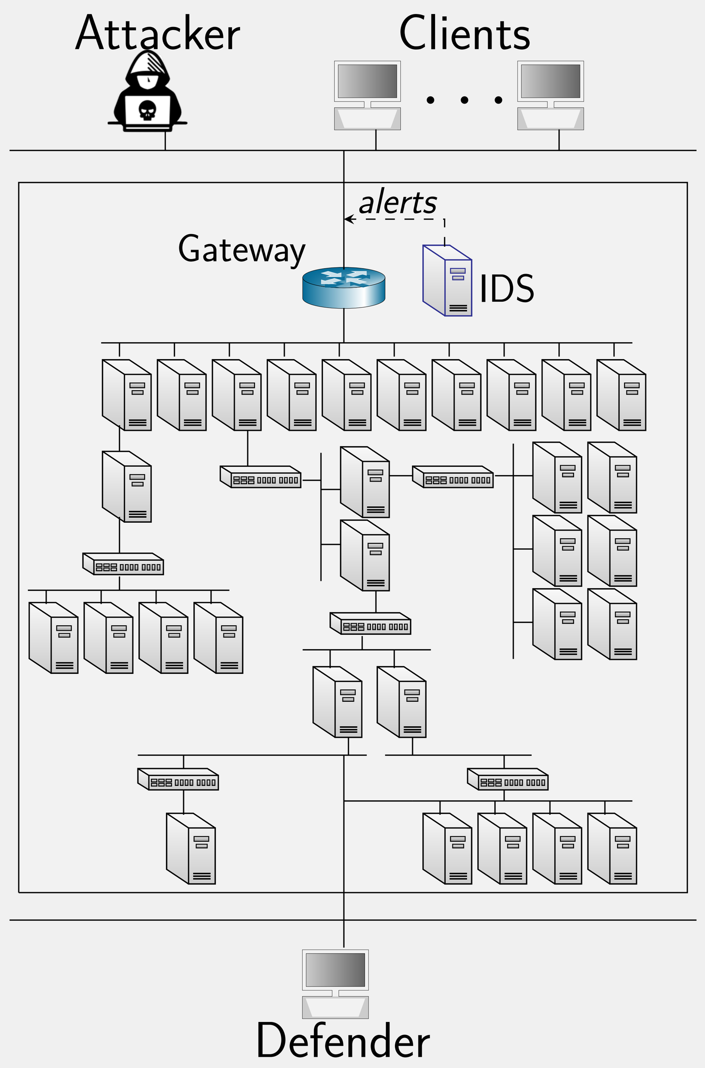
KTH Royal Institute of Technology, EECS, Division of Network and Systems Engineering

Motivation

- **Problem**: Cyber attacks evolve quickly. As a consequence, a defender must constantly adapt and improve the target system to remain effective.
- **Approach**
  We formulate intrusion prevention as a multiple stopping problem and use reinforcement learning to automatically find optimal policies.
- **Contributions**
  1. A novel formulation of the use case as a multiple stopping problem.
  2. A reinforcement learning approach to obtain policies in an emulated infrastructure.

Use Case: Intrusion Prevention

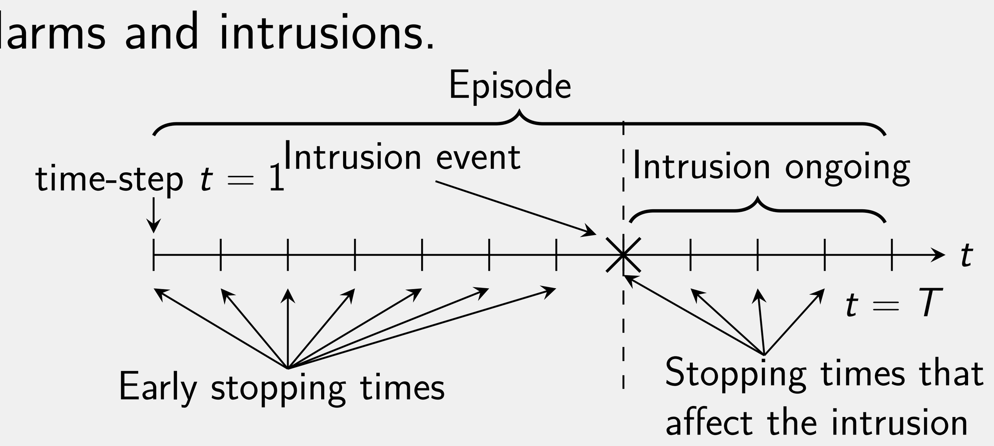
A defender takes measures to protect an IT infrastructure against an attacker while, at the same time, providing a service to a client population.



POMDP Model of the Intrusion Prevention Use Case

We formulate the use case as a **multiple stopping problem** where each stop is associated with a defensive action. We use the following POMDP model:

- **States $\mathcal{S}$ and Observations $\mathcal{O}$:**
  intrusion state $i_t \in \{0,1\}$, $i_t = 1$,
  defender observations $o_t = (\Delta x_t, \Delta y_t, \Delta z_t)$ (IDS alerts and logins).
- **Actions $\mathcal{A}$:** "stop" (S) and "continue" (C)
- **Transition Probabilities $\mathcal{P}^a_{ss'}$ and Observation Function $\mathcal{Z}(o', s', a)$:**
  Intrusion start $(Q_t)_{t=1}^T \sim Ber(p)$.
  Observation distribution $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t, I_t, t)$.
- **Reward Function $\mathcal{R}^a_s$:** Reward for service and intrusion prevention, loss for false alarms and intrusions.

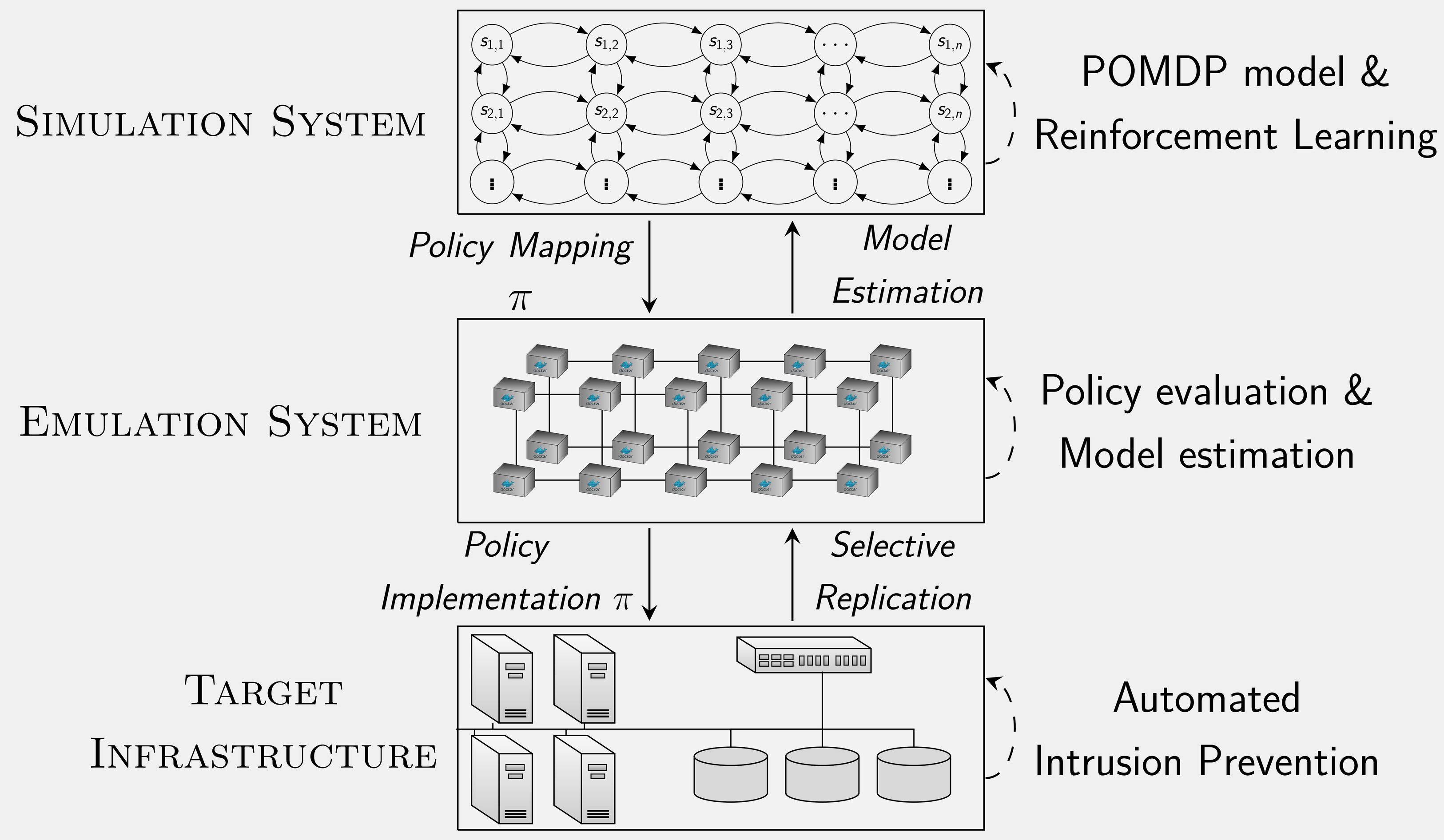


References

- Kim Hammar and Rolf Stadler 2021 ***Intrusion Prevention through Optimal Stopping***. Submitted for publication: https://arxiv.org/abs/2111.00289.
- Kim Hammar and Rolf Stadler 2021 ***Learning Intrusion Prevention Policies through Optimal Stopping***. CNSM 2021. https://ieeexplore.ieee.org/document/9615542
- Kim Hammar and Rolf Stadler 2020 ***Finding Effective Security Strategies through Reinforcement Learning and Self-Play***. CNSM 2020. https://ieeexplore.ieee.org/document/9269092

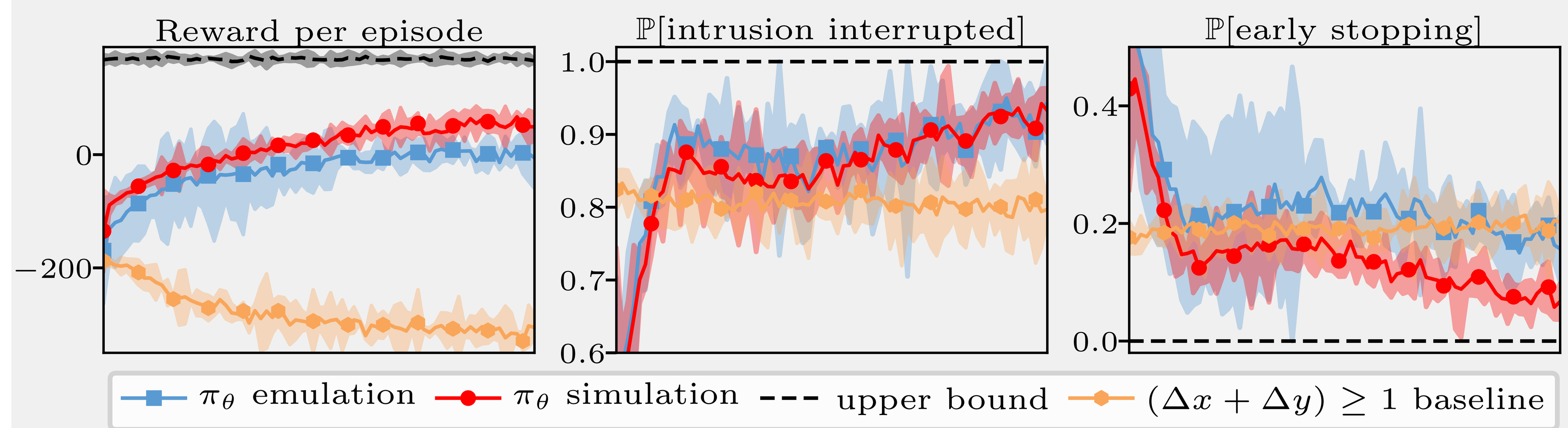
Our Approach

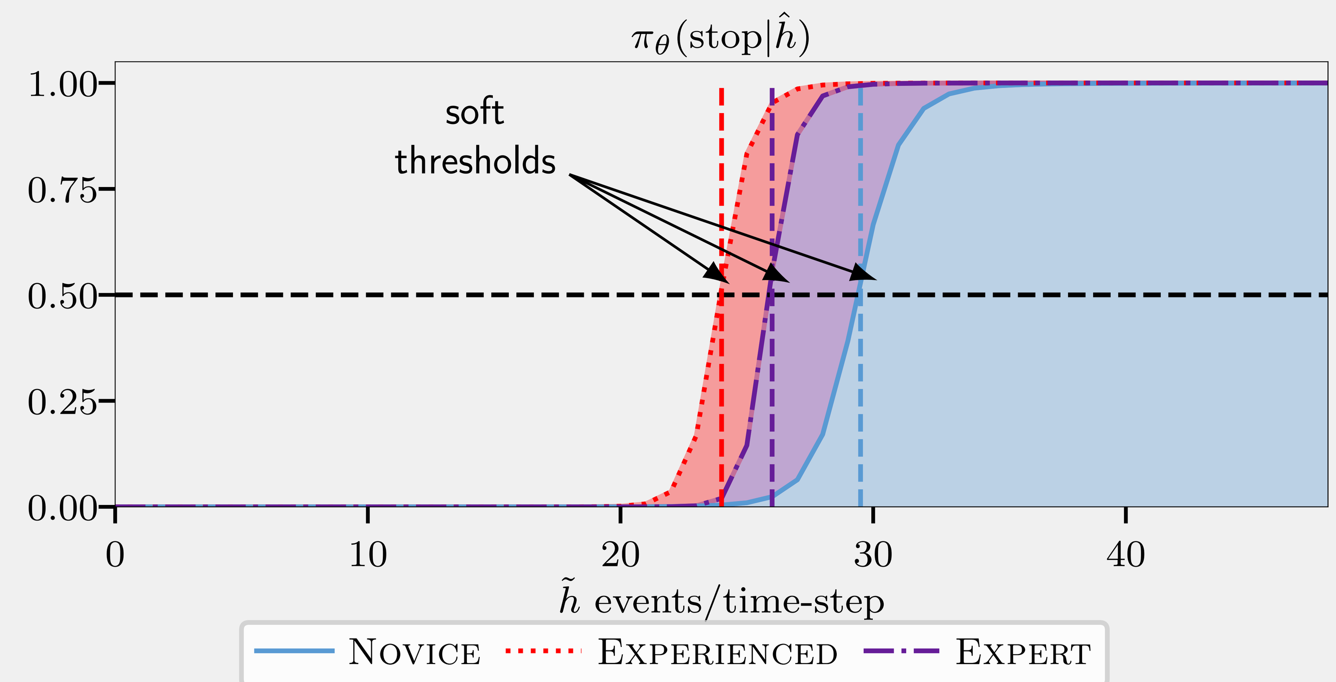
- **The emulation system** replicates key components of the target infrastructure and is used for data collection and policy evaluation.
- **The simulation system** is used to execute POMDP episodes and learn policies through reinforcement learning.



Learning Intrusion Prevention Policies

We use PPO to learn a policy $\pi_\theta : \mathcal{H} \mapsto \mathcal{A}$, where π_θ is a feed-forward neural network and $\mathcal{H}$ is the set of histories.





Threshold Properties of an Optimal Policy

Theorem 1. Let $\mathscr{S}^l$ be the stopping set, and $\mathscr{C}^l$ the continuation set. The following holds:

(A) $\mathscr{S}^{l-1} \subseteq \mathscr{S}^l$ for $l = 2, \ldots L$.

(B) If $L - l^A = 1$, there exists $\alpha^* \in [0,1]$ and an optimal policy π^*_l that satisfies:

$$\pi^*_L(b(1)) = S \iff b(1) \geq \alpha^* \qquad (1)$$

(C) If $L - l^A \geq 1$ and $f_{XYZ|s}$ is totally positive of order 2 (i.e., TP2), there exist $L - l^A$ values $\alpha^*_{l^A+1} \geq \alpha^*_{l^A+2} \geq \ldots \geq \alpha^*_L \in [0,1]$ and an optimal policy π^*_l that satisfies:

$$\pi^*_l(b(1)) = S \iff b(1) \geq \alpha^*_l, l \in l^A + 1, \ldots, L \qquad (2)$$

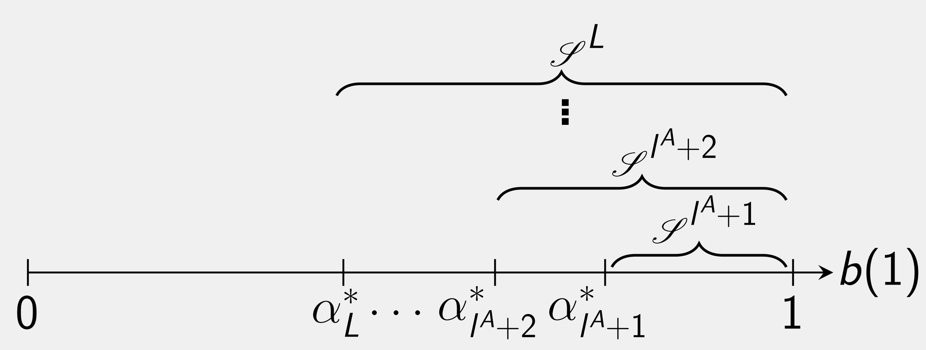


Figure: Illustration of Theorem 1: there exist $L - l^A$ thresholds $\alpha^*_{l^A+1} \geq \alpha^*_{l^A+2} \ldots, \geq \alpha^*_L \in \mathcal{B}$ and an optimal threshold policy π^*_l.