Learning Near-Optimal Intrusion Responses Against Dynamic Attackers

CDIS Spring Conference, 25 May 2023

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

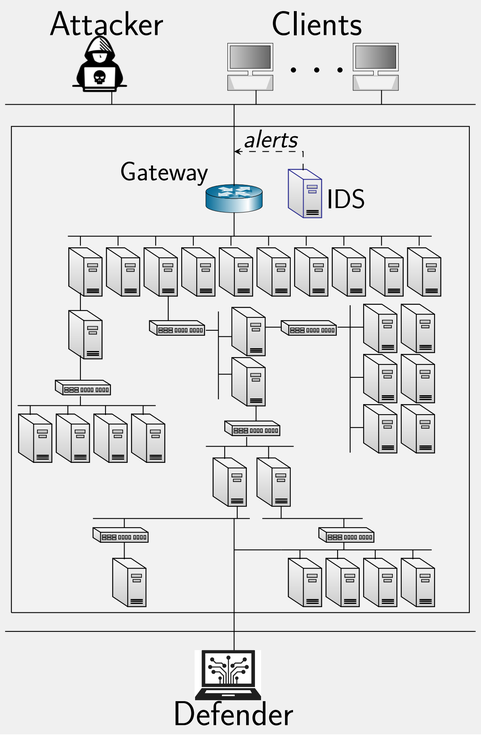
KTH Royal Institute of Technology, Center for Cyber Defense and Information Security (CDIS)

Motivation and Contributions

- **Problem**: Cyber attacks evolve quickly. As a consequence, a defender must constantly adapt and improve the target system to remain effective.
- **Contributions**
  1. A novel formulation of intrusion response as an optimal stopping game.
  2. A method to obtain strategies with demonstrated performance in emulated infrastructures.
  3. A reinforcement learning algorithm (T-FP) that outperforms state-of-the-art.

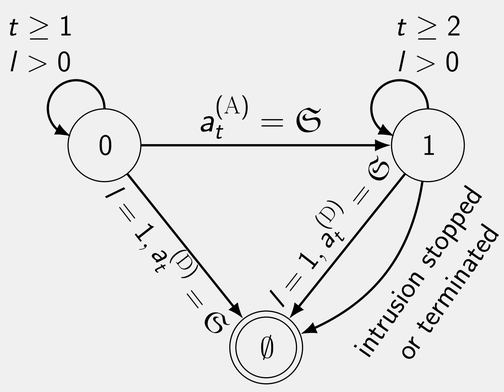
Use Case: Intrusion Response

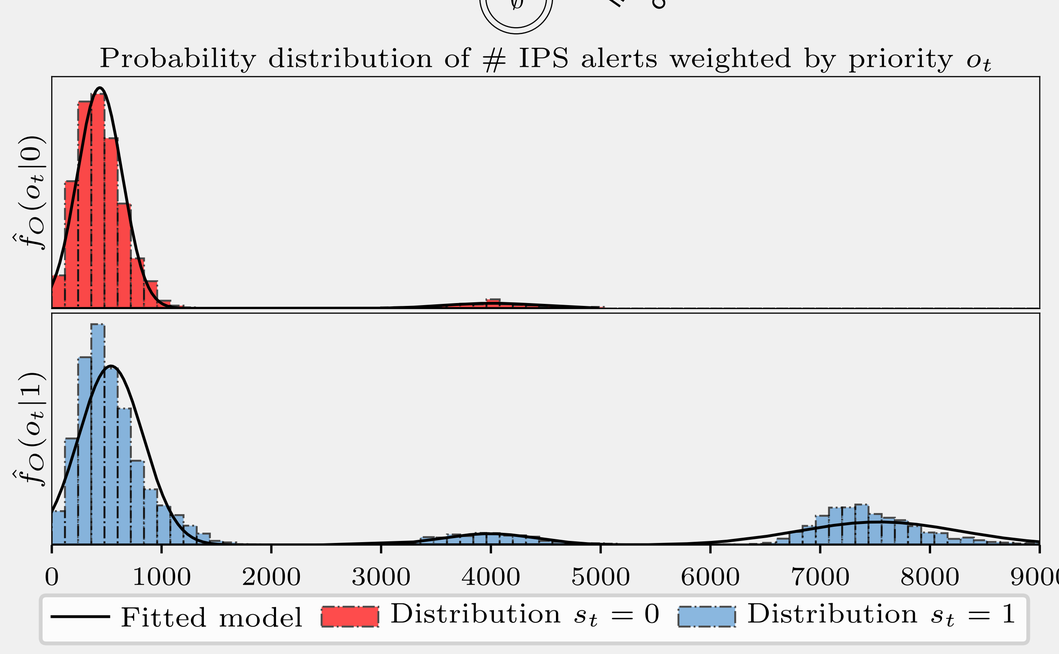
A defender takes measures to protect an IT infrastructure against an attacker while providing services to a client population.



Partially Observed Stochastic Stopping Game

We formulate the use case as a **Partially Observed Stochastic stopping Game (POSG)**. Each stop action of the defender correspond to a measure against a possible intrusion. The attackers' stop actions determine when the intrusion starts and stops.

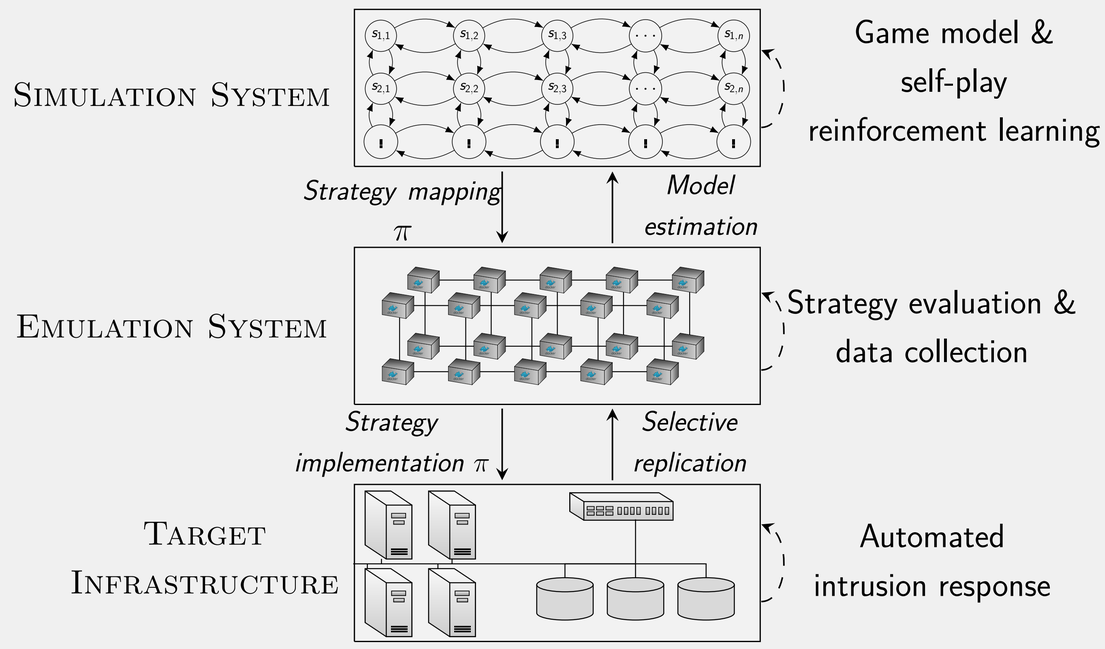




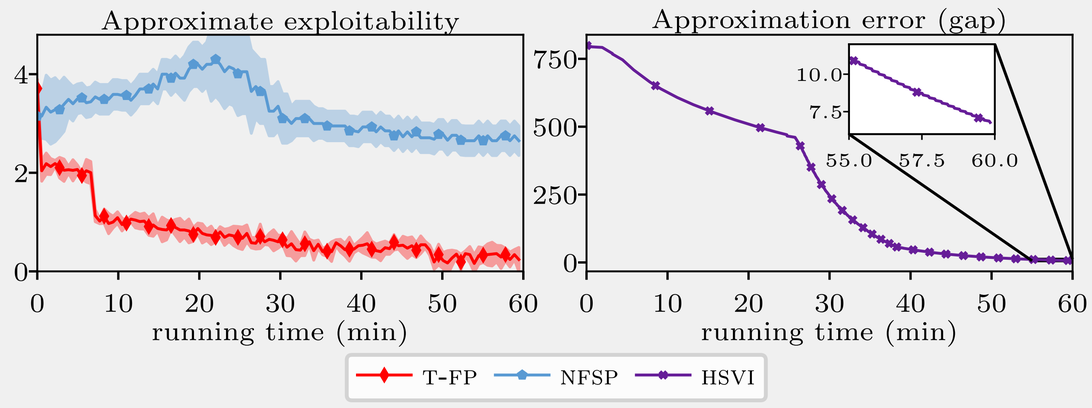
Video of Software Framework

Our Approach

- **The emulation system** replicates key components of the target infrastructure and is used for data collection and strategy evaluation.
- **The simulation system** is used to simulate game episodes and learn strategies through reinforcement learning.



Learning Optimal (Equilibrium) Strategies with T-FP



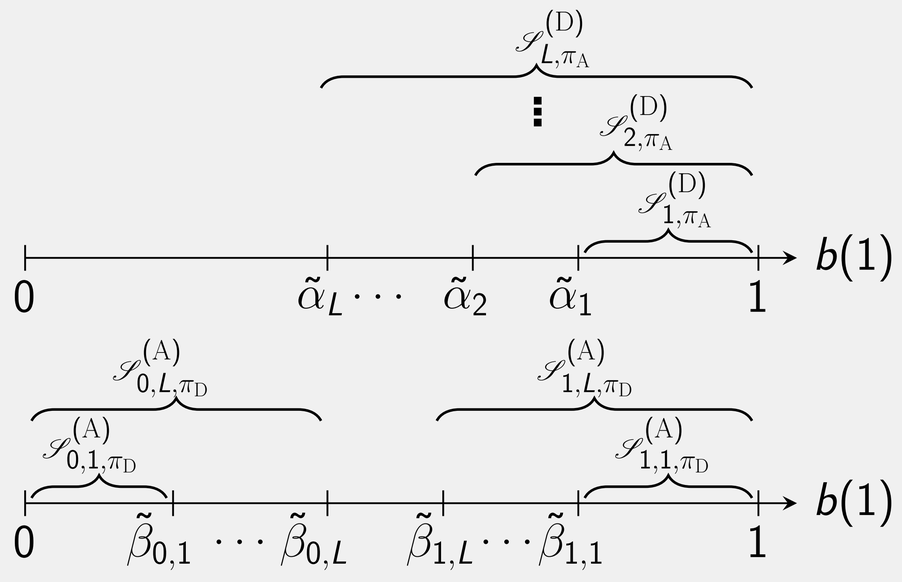
Threshold Properties of Best Response Strategies

Theorem 1. Given the POSG Γ with one-sided partial observability and $L \geq 1$ stop actions for the defender, the following holds.

(A) Γ has a mixed Nash equilibrium. If $s = 0 \iff b(1) = 0$, then it has a pure Nash equilibrium.

(B) If $f_{O|s}$ is totally positive of order 2, there exist L values $\tilde{\alpha}_1 \geq \tilde{\alpha}_2 \geq \ldots \geq \tilde{\alpha}_L \in [0, 1]$ and a best response multi-threshold defender strategy $\tilde{\pi}_D$.

(C) If the π_D is non-decreasing in $b(1)$, then there exist values $\tilde{\beta}_{0,1}$, $\tilde{\beta}_{1,1}$, ..., $\tilde{\beta}_{0,L}$, $\tilde{\beta}_{1,L} \in [0, 1]$ and a best response multi-threshold attacker strategy $\tilde{\pi}_A$.



References

- Kim Hammar and Rolf Stadler 2023 ***Learning Near-Optimal Intrusion Responses Against Dynamic Attackers***. To appear in IEEE TNSM. https://arxiv.org/abs/2301.06085.
- Kim Hammar and Rolf Stadler 2022 ***Intrusion Prevention through Optimal Stopping***. IEEE TNSM. https://ieeexplore.ieee.org/document/9779345.