Intrusion Prevention through Optimal Stopping

CDIS Spring Conference, 24 May 2022

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

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

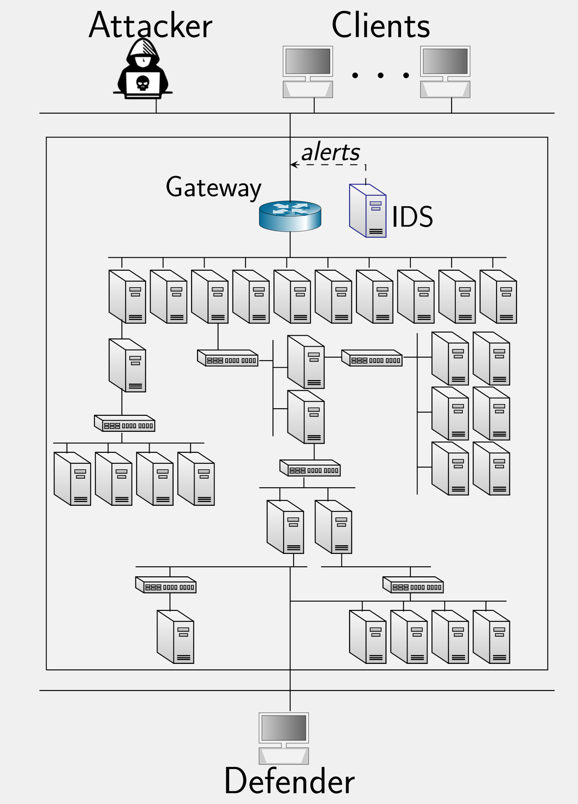
Motivation and Contributions

- **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 prevention as a multiple stopping problem.
  2. A method to obtain policies with demonstrated performance in emulated infrastructures.
  3. A reinforcement learning algorithm (T-SPSA) that outperforms state-of-the-art.

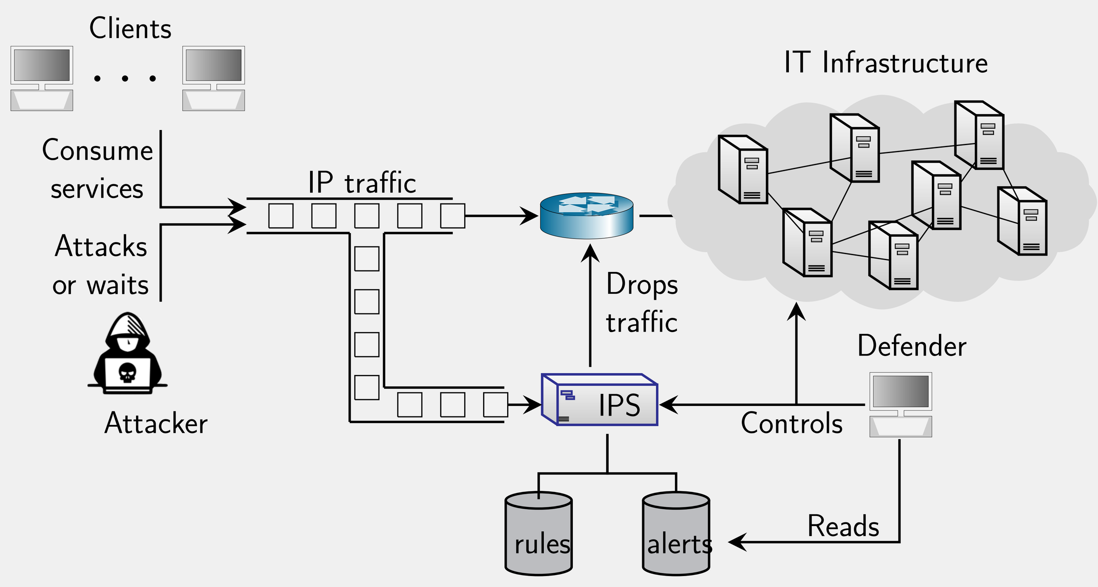
Use Case: Intrusion Prevention

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

a) The infrastructure and the actors in the use case.

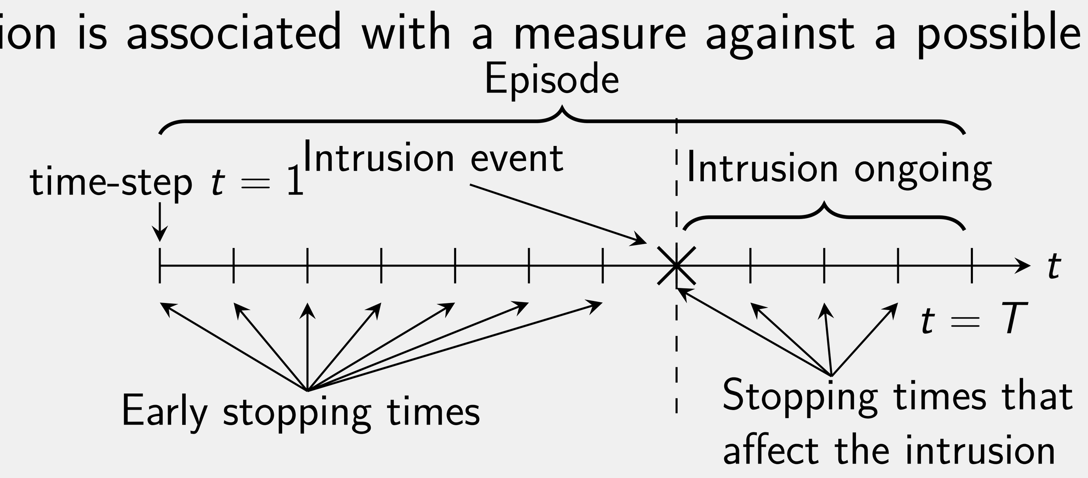


b) The game between the attacker and the defender

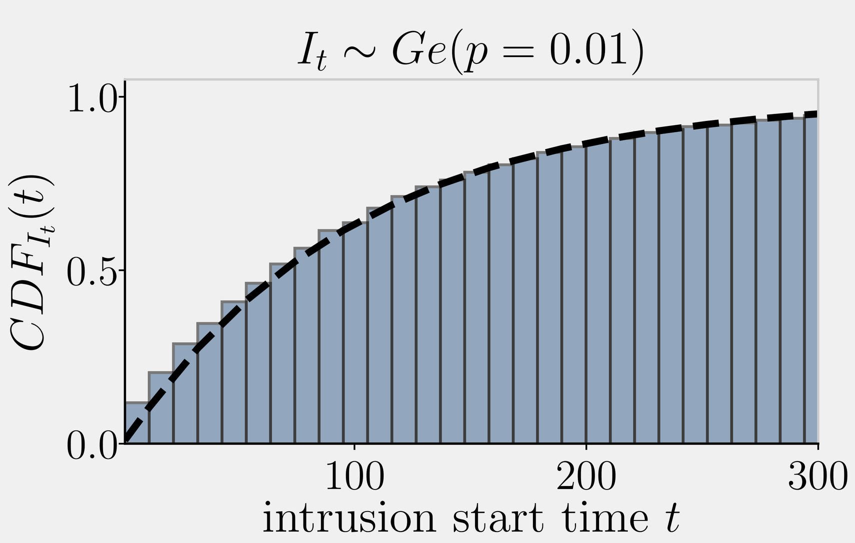


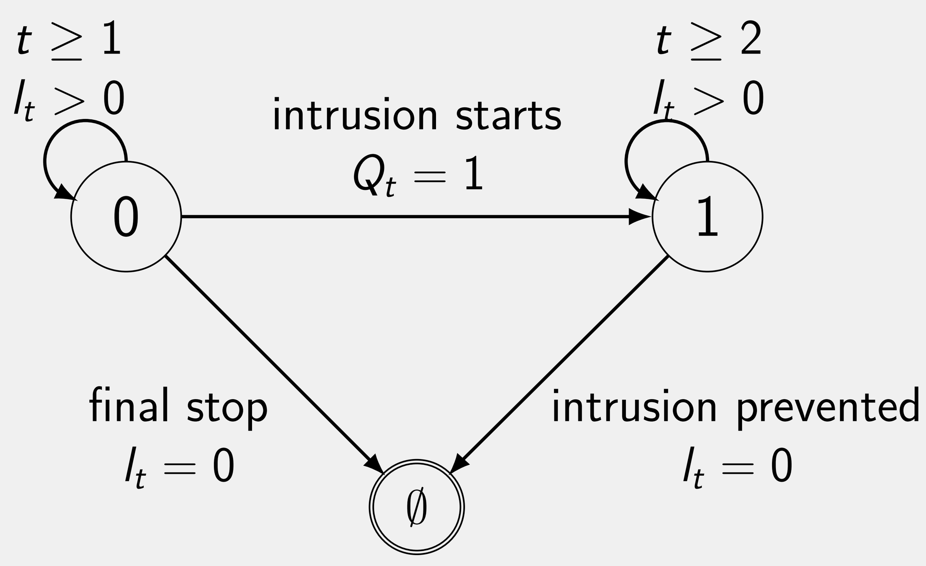
POMDP Model of the Intrusion Prevention Use Case

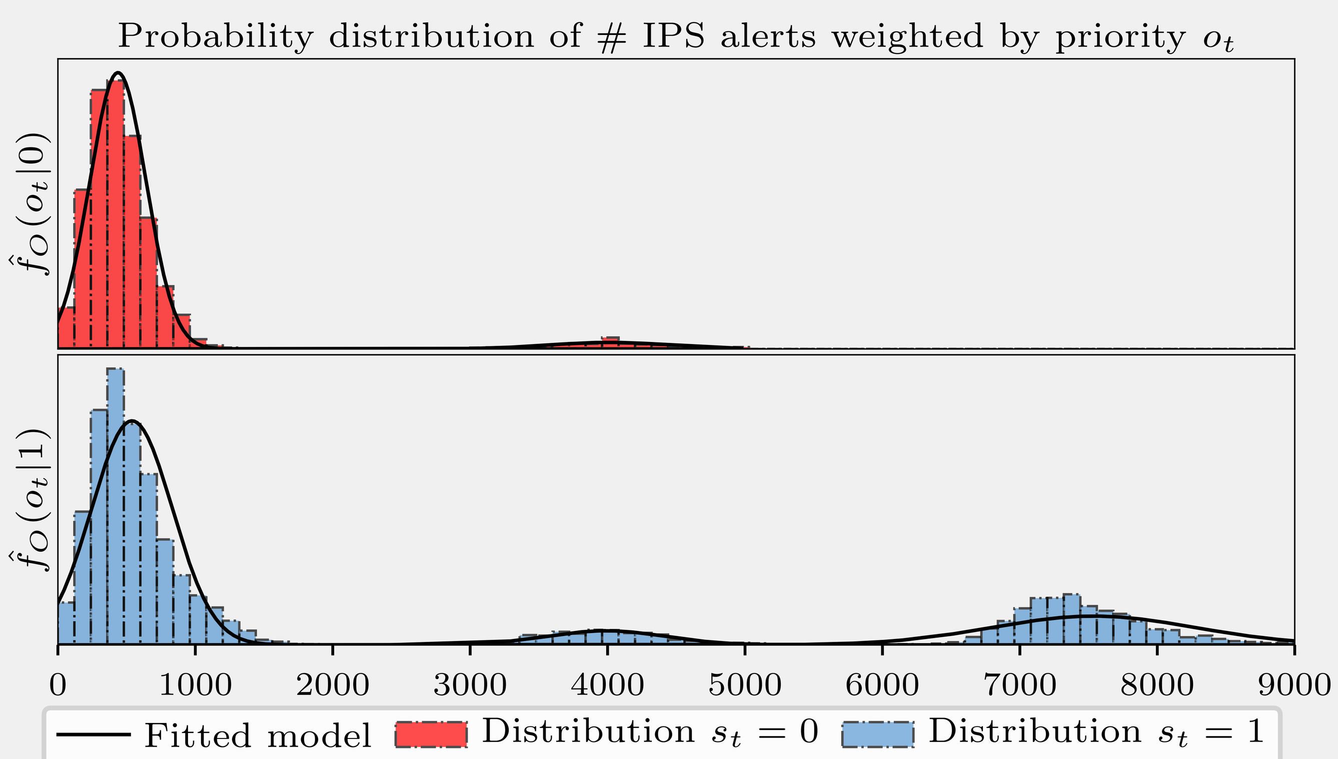
We formulate the use case as a **multiple stopping problem**, where each stopping action is associated with a measure against a possible intrusion.



We use the following POMDP model:



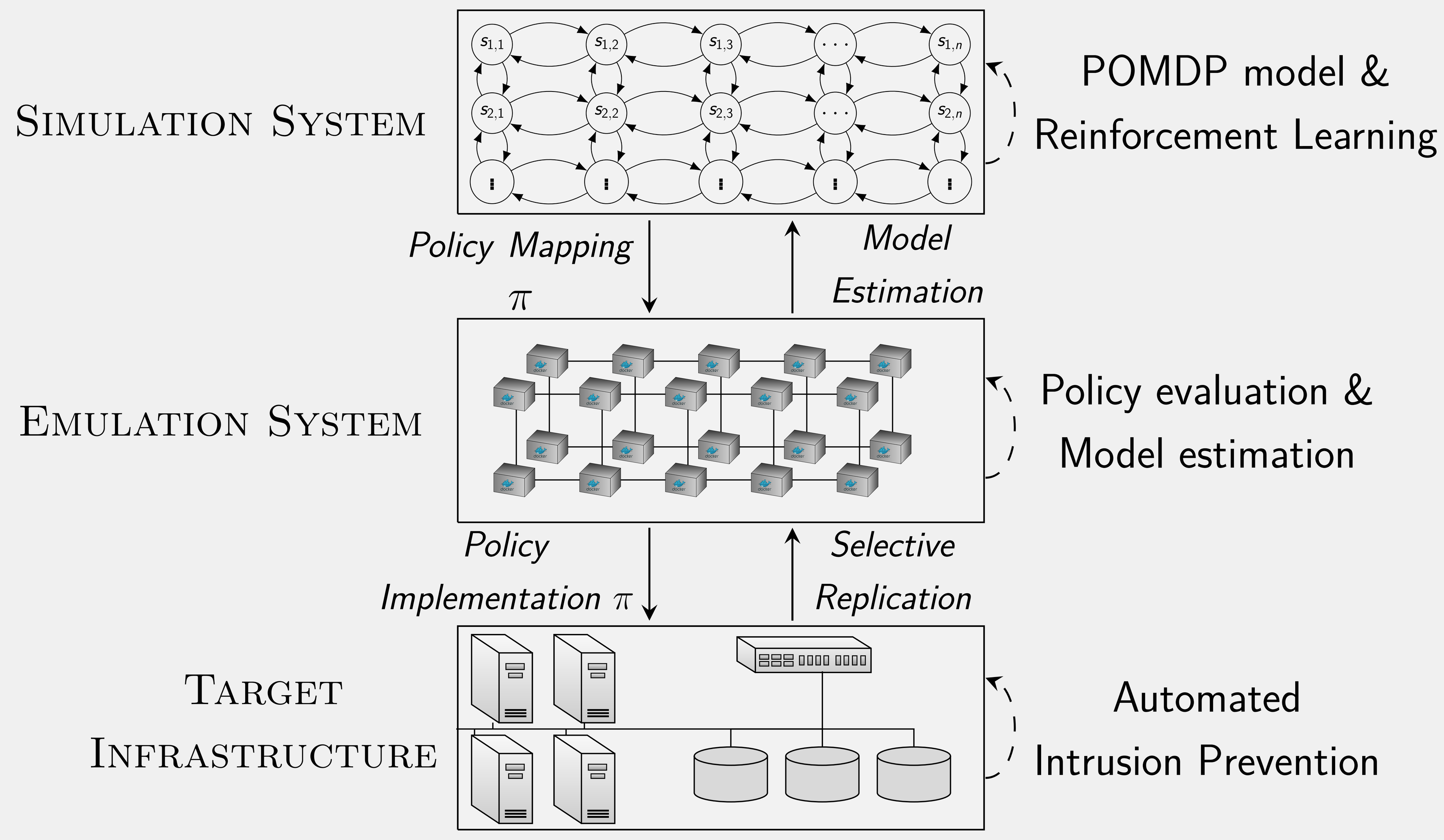




Video of Software Framework

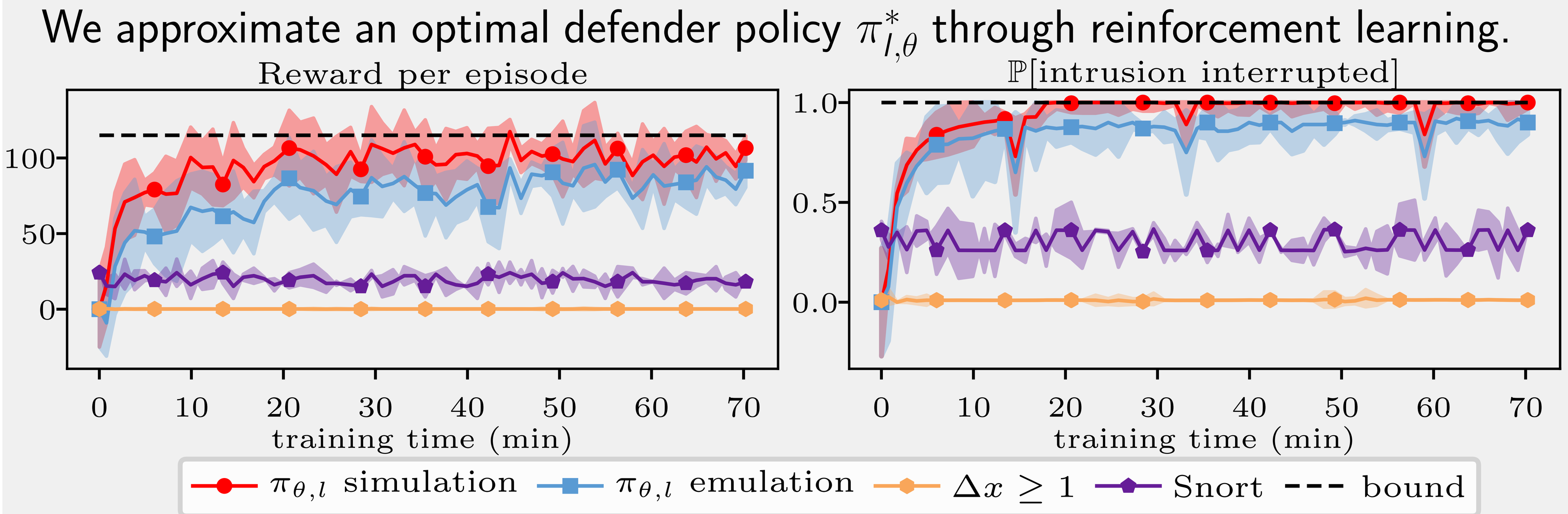
Our Approach

- **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 with T-SPSA

We approximate an optimal defender policy $\pi^*_{l,\theta}$ through reinforcement learning.



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 $l = 2, \dots L$.

(B) If $L = 1$, there exists a value $\alpha^* \in [0,1]$ and an optimal policy π^*_L that satisfies:

$$\pi^*_L(b(1)) = S \iff b(1) \geq \alpha^* \tag{1}$$

(C) If $L \geq 1$ and $f_{XYZ|s}$ is totally positive of order 2 (i.e., TP2), there exist L values $\alpha^*_1 \geq \alpha^*_2 \geq \dots \geq \alpha^*_L \in [0,1]$ and an optimal policy π^*_l that satisfies:

$$\pi^*_l(b(1)) = S \iff b(1) \geq \alpha^*_l \quad l \in \{1,\dots,L\} \tag{2}$$

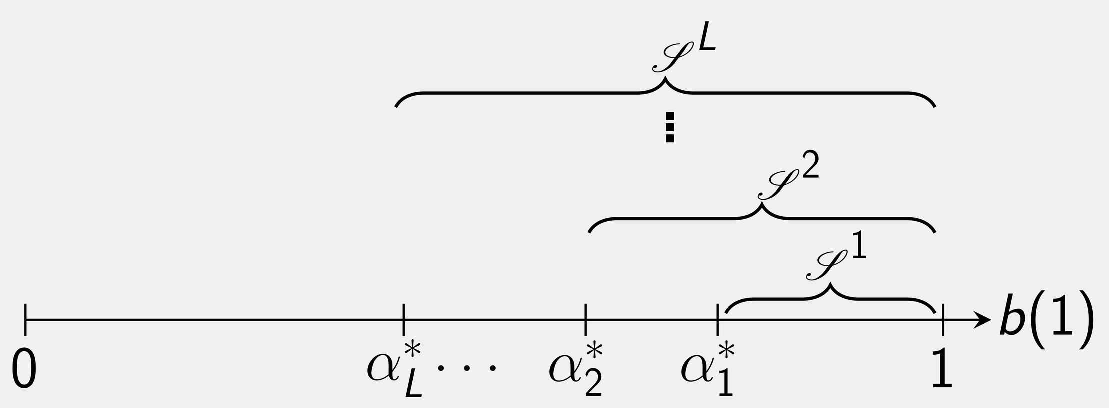


Figure: Illustration of Theorem 1.

References

- Kim Hammar and Rolf Stadler 2022 ***A System for Interactive Examination of Learned Security Policies***. NOMS 2022. https://arxiv.org/abs/2204.01126.
- Kim Hammar and Rolf Stadler 2021 ***Intrusion Prevention through Optimal Stopping***. To appear in IEEE TNSM. 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