Intrusion Prevention through Optimal Stopping

Netcon talk

Kim Hammar & Rolf Stadler

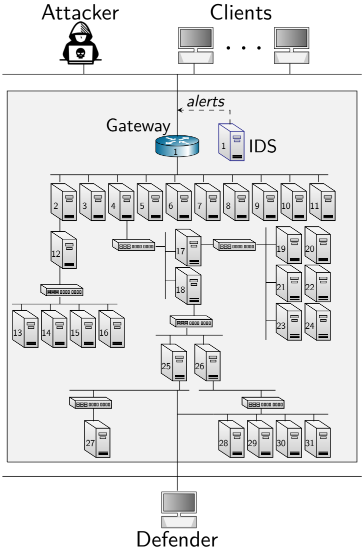
kimham@kth.se & *stadler@kth.se*

Division of Network and Systems Engineering
KTH Royal Institute of Technology

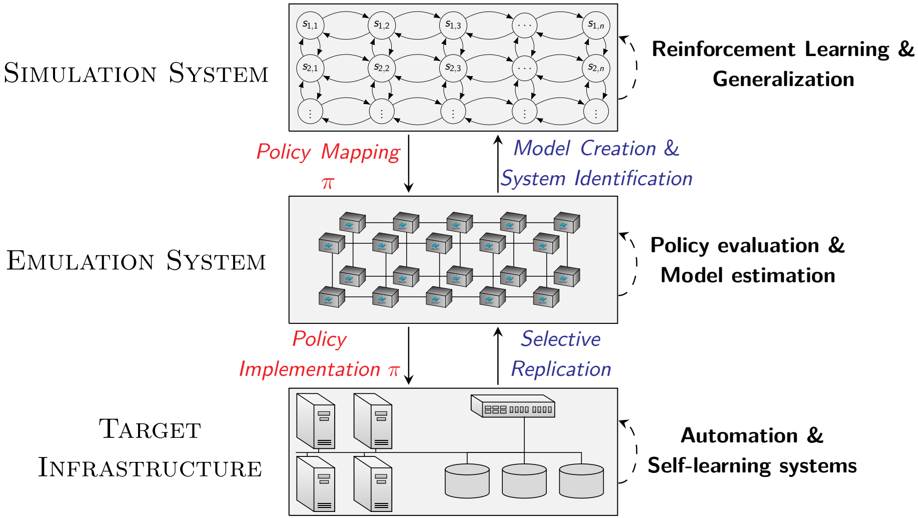
Feb 7, 2022

Use Case: Intrusion Prevention

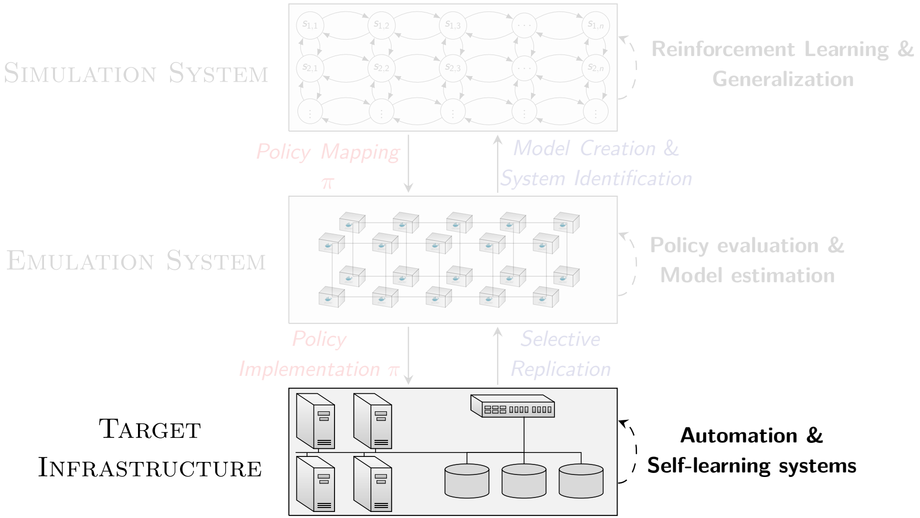
- A **Defender** owns an infrastructure
  - Consists of connected components
  - Components run network services
  - Defender defends the infrastructure by monitoring and active defense
- An **Attacker** seeks to intrude on the infrastructure
  - Has a partial view of the infrastructure
  - Wants to compromise specific components
  - Attacks by reconnaissance, exploitation and pivoting



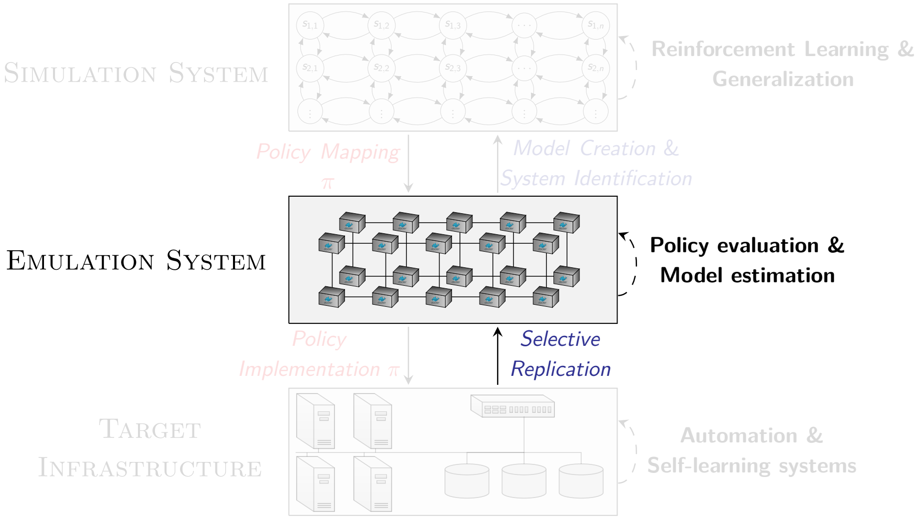
Our Method for Finding Effective Security Strategies



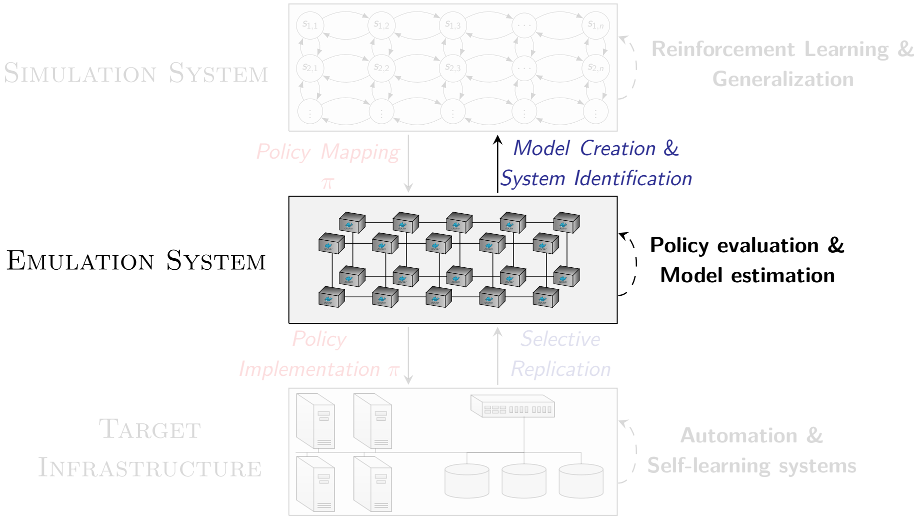
Our Method for Finding Effective Security Strategies



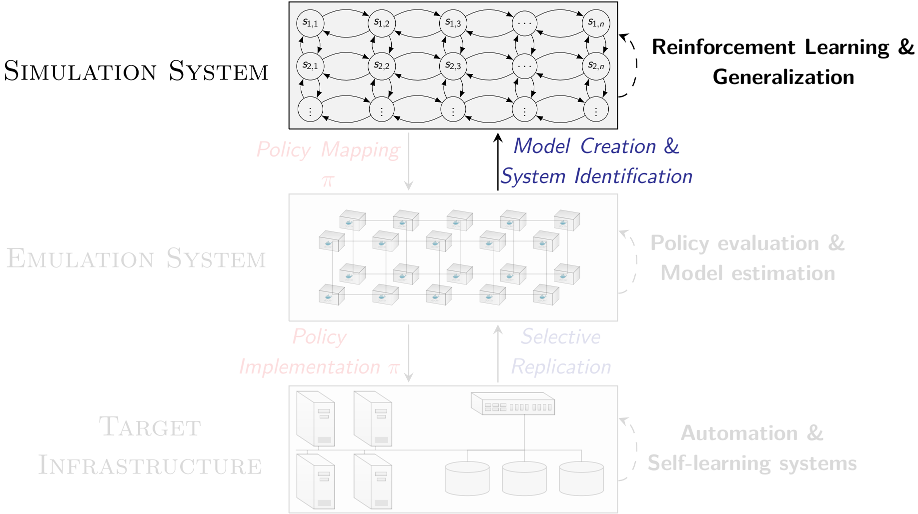
Our Method for Finding Effective Security Strategies



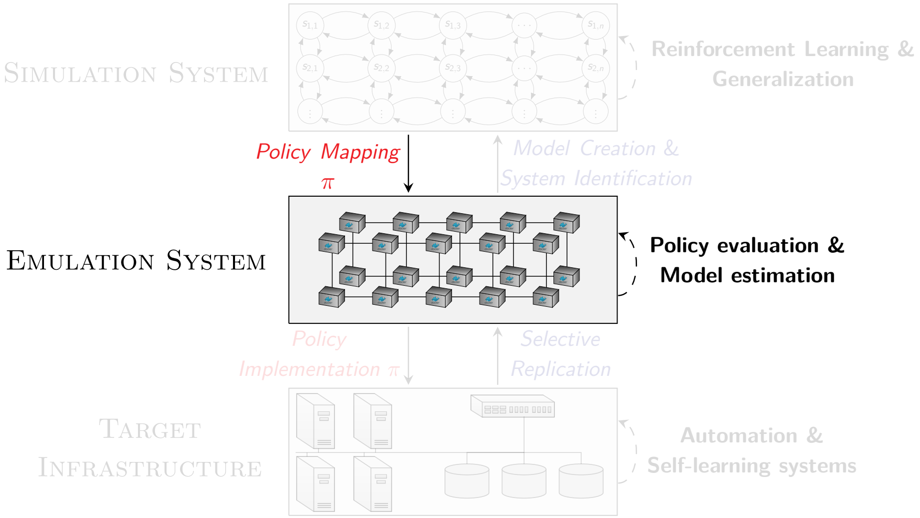
Our Method for Finding Effective Security Strategies



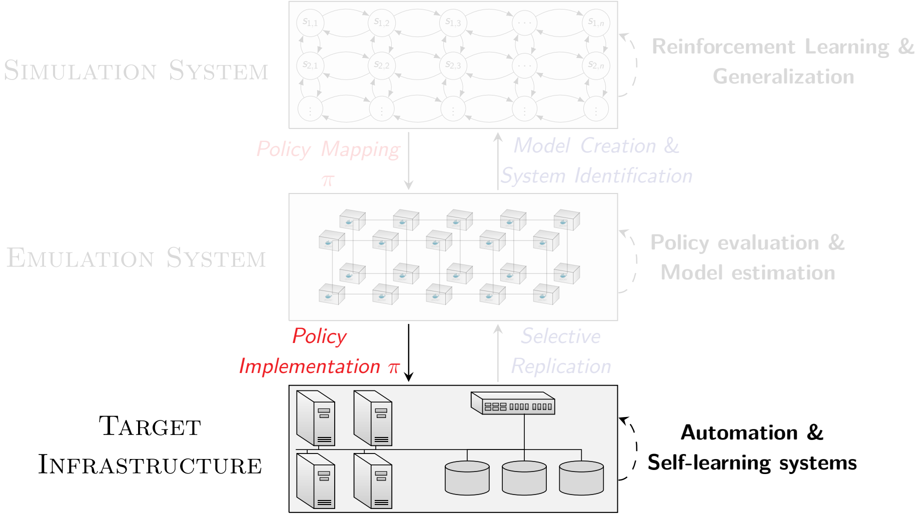
Our Method for Finding Effective Security Strategies



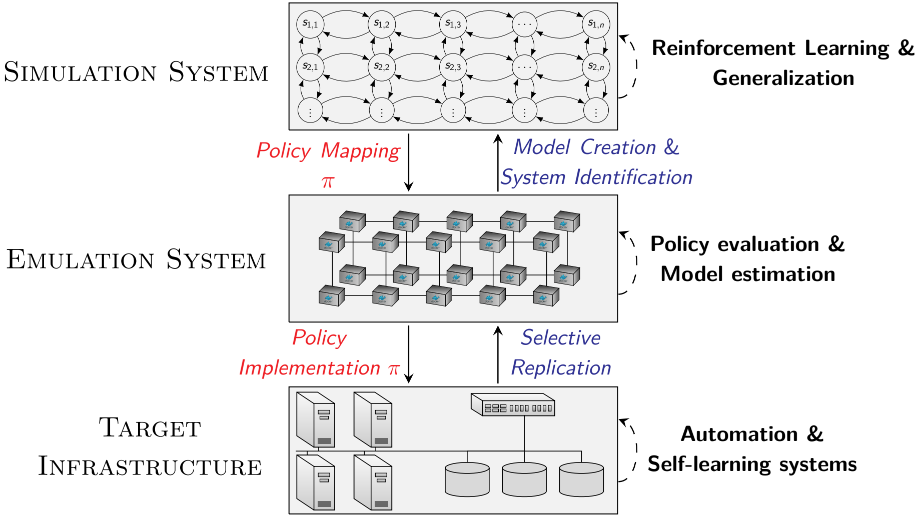
Our Method for Finding Effective Security Strategies



Our Method for Finding Effective Security Strategies



Our Method for Finding Effective Security Strategies



Outline

- **Use Case & Approach:**
  - Intrusion Prevention
  - System identification
  - Reinforcement learning and dynamic programming
- **Formal Model & Background:**
  - Background: POMDPs and optimal stopping
  - Multiple Stopping Problem POMDP
- **Structure of π^***
  - Structural result: Multi-Threshold policy
  - Stopping sets $\mathscr{S}_l$ are connected and nested
  - Conditions for Bayesian filter to be monotone in b
  - Existence of optimal multi-threshold policy π_l^*
- **Conclusion**
  - Numerical evaluation results
  - Conclusion & Future work

Outline

- **Use Case & Approach:**
  - Intrusion Prevention
  - System identification
  - Reinforcement learning and dynamic programming
- **Formal Model & Background:**
  - Background: POMDPs and optimal stopping
  - Multiple Stopping Problem POMDP
- **Structure of π^***
  - Structural result: Multi-Threshold policy
  - Stopping sets $\mathscr{S}_l$ are connected and nested
  - Conditions for Bayesian filter to be monotone in b
  - Existence of optimal multi-threshold policy π^*_l
- **Conclusion**
  - Numerical evaluation results
  - Conclusion & Future work

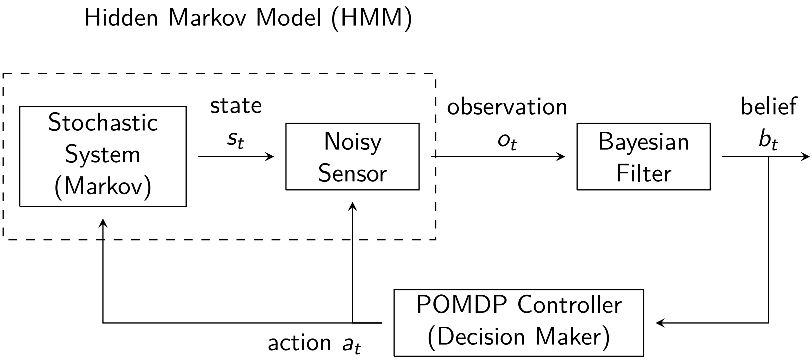
Outline

- **Use Case & Approach:**
  - Intrusion Prevention
  - System identification
  - Reinforcement learning and dynamic programming
- **Formal Model & Background:**
  - Background: POMDPs and optimal stopping
  - Multiple Stopping Problem POMDP
- **Structure of π^***
  - Structural result: Multi-Threshold policy
  - Stopping sets $\mathscr{S}_l$ are connected and nested
  - Conditions for Bayesian filter to be monotone in b
  - Existence of optimal multi-threshold policy π^*_l
- **Conclusion**
  - Numerical evaluation results
  - Conclusion & Future work

Outline

- **Use Case & Approach:**
  - Intrusion Prevention
  - System identification
  - Reinforcement learning and dynamic programming
- **Formal Model & Background:**
  - Background: POMDPs and optimal stopping
  - Multiple Stopping Problem POMDP
- **Structure of π^***
  - Structural result: Multi-Threshold policy
  - Stopping sets $\mathscr{S}_l$ are connected and nested
  - Conditions for Bayesian filter to be monotone in b
  - Existence of optimal multi-threshold policy π_l^*
- **Conclusion**
  - Numerical evaluation results
  - Conclusion & Future work

Background: POMDPs



- **POMDP:** $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}^{a_t}_{s_t,s_{t+1}}, \mathcal{R}^{a_t}_{s_t,s_{t+1}}, \gamma, \rho_1, T, \mathcal{O}, \mathcal{Z} \rangle$
- Controlled hidden Markov model, states $s_t \in \mathcal{S}$ are hidden
- Agent observes history $h_t = (\rho_1, a_1, o_1, \ldots, a_{t-1}, o_t) \in \mathcal{H}$

Background: POMDPs

- s_t is Markov: $\mathbb{P}[s_{t+1}|s_t] = \mathbb{P}[s_{t+1}|s_1, \ldots, s_t]$
- $\implies \pi^*(a_t|h_t) = \pi^*(a_t|\mathbb{P}[s_t|h_t]) = \pi^*(a_t|b_t)$
- **Optimality (Bellman) Eq:**

$$\pi^*(b) \in \arg\max_{a \in \mathcal{A}} \left[\sum_s b(s)\mathcal{R}_s^a + \gamma \sum_{o,s,s'} \mathcal{Z}(o,s',a) b(s) \mathcal{P}_{ss'}^a V^*(b_a^o) \right]$$

-

$$\begin{aligned}
\mathbb{P}[s_t|h_t] &= \mathbb{P}[s_t|o_t, a_{t-1}, h_{t-1}] \\
&= \frac{\mathbb{P}[o_t|s_t, a_{t-1}, h_{t-1}]\mathbb{P}[s_t|a_{t-1}, h_{t-1}]}{\mathbb{P}[o_t|a_{t-1}, h_{t-1}]} && \text{Bayes} \\
&= \frac{\mathcal{Z}(o_t, s_t, a_{t-1}) \sum_{s_{t-1}} \mathcal{P}_{s_{t-1}s_t}^{a_{t-1}} \mathbb{P}[s_{t-1}|h_{t-1}]}{\sum_{s'} \sum_s \mathcal{Z}(o_t, s', a_{t-1}) \mathbb{P}[s_{t-1}|h_{t-1}]} && \text{Markov}
\end{aligned}$$

- $\mathbb{P}[s_{t-1}|h_{t-1}]$ **with** a_t, o_t **is a sufficient statistic for** s_t
- $b_t \triangleq \mathbb{P}[s_{t-1}|h_{t-1}]$**: belief state at time** t
- b_t **computed recursively using the equation above**

Background: POMDPs

- s_t is Markov: $\mathbb{P}[s_{t+1}|s_t] = \mathbb{P}[s_{t+1}|s_1, \ldots, s_t]$
- $\implies \pi^*(a_t|h_t) = \pi^*(a_t|\mathbb{P}[s_t|h_t]) = \pi^*(a_t|b_t)$
- **Optimality (Bellman) Eq:**

$$\pi^*(b) \in \arg\max_{a\in\mathcal{A}} \left[\sum_s b(s)\mathcal{R}_s^a + \gamma \sum_{o,s,s'} \mathcal{Z}(o,s',a)b(s)\mathcal{P}_{ss'}^a V^*(b_a^o) \right]$$

-

$$\begin{aligned}
\mathbb{P}[s_t|h_t] &= \mathbb{P}[s_t|o_t, a_{t-1}, h_{t-1}] \\
&= \frac{\mathbb{P}[o_t|s_t, a_{t-1}, h_{t-1}]\mathbb{P}[s_t|a_{t-1}, h_{t-1}]}{\mathbb{P}[o_t|a_{t-1}, h_{t-1}]} && \text{Bayes} \\
&= \frac{\mathcal{Z}(o_t, s_t, a_{t-1}) \sum_{s_{t-1}} \mathcal{P}_{s_{t-1}s_t}^{a_{t-1}} \mathbb{P}[s_{t-1}|h_{t-1}]}{\sum_{s'} \sum_s \mathcal{Z}(o_t, s', a_{t-1}) \mathbb{P}[s_{t-1}|h_{t-1}]} && \text{Markov}
\end{aligned}$$

- $\mathbb{P}[s_{t-1}|h_{t-1}]$ **with** a_t, o_t **is a sufficient statistic for** s_t
- $b_t \triangleq \mathbb{P}[s_{t-1}|h_{t-1}]$**: belief state at time** t
- b_t **computed recursively using the equation above**

Background: POMDPs

- s_t is Markov: $\mathbb{P}\left[s_{t+1}|s_t\right] = \mathbb{P}\left[s_{t+1}|s_1,\ldots,s_t\right]$
- $\implies \pi^*(a_t|h_t) = \pi^*(a_t|\mathbb{P}[s_t|h_t]) = \pi^*(a_t|b_t)$
- **Optimality (Bellman) Eq:**

$$\pi^*(b) \in \arg\max_{a\in\mathcal{A}} \left[\sum_s b(s)\mathcal{R}_s^a + \gamma \sum_{o,s,s'} \mathcal{Z}(o,s',a)b(s)\mathcal{P}_{ss'}^a V^*(b_a^o)\right]$$

-

$$\begin{aligned}
\mathbb{P}[s_t|h_t] &= \mathbb{P}[s_t|o_t,a_{t-1},h_{t-1}] \\
&= \frac{\mathbb{P}[o_t|s_t,a_{t-1},h_{t-1}]\mathbb{P}[s_t|a_{t-1},h_{t-1}]}{\mathbb{P}[o_t|a_{t-1},h_{t-1}]} && \text{Bayes} \\
&= \frac{\mathcal{Z}(o_t,s_t,a_{t-1})\sum_{s_{t-1}}\mathcal{P}_{s_{t-1}s_t}^{a_{t-1}}\mathbb{P}[s_{t-1}|h_{t-1}]}{\sum_{s'}\sum_s \mathcal{Z}(o_t,s',a_{t-1})\mathbb{P}[s_{t-1}|h_{t-1}]} && \text{Markov}
\end{aligned}$$

- $\mathbb{P}[s_{t-1}|h_{t-1}]$ **with** a_t, o_t **is a sufficient statistic for** s_t
- $b_t \triangleq \mathbb{P}[s_{t-1}|h_{t-1}]$: **belief state at time** t
- b_t **computed recursively using the equation above**

Background: POMDPs

- s_t is Markov: $\mathbb{P}[s_{t+1}|s_t] = \mathbb{P}[s_{t+1}|s_1,\dots,s_t]$
- $\implies \pi^*(a_t|h_t) = \pi^*(a_t|\mathbb{P}[s_t|h_t]) = \pi^*(a_t|b_t)$
- **Optimality (Bellman) Eq:**

$$\pi^*(b) \in \arg\max_{a\in\mathcal{A}} \left[\sum_s b(s)\mathcal{R}_s^a + \gamma \sum_{o,s,s'} \mathcal{Z}(o,s',a)b(s)\mathcal{P}_{ss'}^a V^*(b_a^o) \right]$$

-

$$\begin{aligned}
\mathbb{P}[s_t|h_t] &= \mathbb{P}[s_t|o_t, a_{t-1}, h_{t-1}] \\
&= \frac{\mathbb{P}[o_t|s_t, a_{t-1}, h_{t-1}]\mathbb{P}[s_t|a_{t-1}, h_{t-1}]}{\mathbb{P}[o_t|a_{t-1}, h_{t-1}]} && \text{Bayes} \\
&= \frac{\mathcal{Z}(o_t, s_t, a_{t-1})\sum_{s_{t-1}} \mathcal{P}_{s_{t-1}s_t}^{a_{t-1}} \mathbb{P}[s_{t-1}|h_{t-1}]}{\sum_{s'}\sum_s \mathcal{Z}(o_t, s', a_{t-1})\mathbb{P}[s_{t-1}|h_{t-1}]} && \text{Markov}
\end{aligned}$$

- $\mathbb{P}[s_{t-1}|h_{t-1}]$ **with** a_t, o_t **is a sufficient statistic for** s_t
- $b_t \triangleq \mathbb{P}[s_{t-1}|h_{t-1}]$**: belief state at time** t
- b_t **computed recursively using the equation above**

Background: POMDPs

- s_t is Markov: $\mathbb{P}\left[s_{t+1}|s_t\right] = \mathbb{P}\left[s_{t+1}|s_1, \ldots, s_t\right]$
- $\implies \pi^*(a_t|h_t) = \pi^*(a_t|\mathbb{P}[s_t|h_t]) = \pi^*(a_t|b_t)$
- **Optimality (Bellman) Eq:**

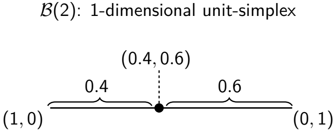
$$\pi^*(b) \in \arg\max_{a \in \mathcal{A}} \left[\sum_s b(s)\mathcal{R}_s^a + \gamma \sum_{o,s,s'} \mathcal{Z}(o,s',a) b(s) \mathcal{P}_{ss'}^a V^*(b_a^o) \right]$$

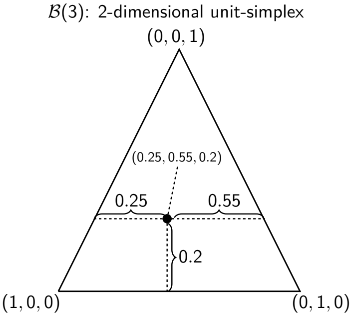
-

$$\begin{aligned}
\mathbb{P}[s_t|h_t] &= \mathbb{P}[s_t|o_t, a_{t-1}, h_{t-1}] \\
&= \frac{\mathbb{P}[o_t|s_t, a_{t-1}, h_{t-1}]\mathbb{P}[s_t|a_{t-1}, h_{t-1}]}{\mathbb{P}[o_t|a_{t-1}, h_{t-1}]} && \text{Bayes} \\
&= \frac{\mathcal{Z}(o_t, s_t, a_{t-1}) \sum_{s_{t-1}} \mathcal{P}_{s_{t-1}s_t}^{a_{t-1}} \mathbb{P}[s_{t-1}|h_{t-1}]}{\sum_{s'} \sum_s \mathcal{Z}(o_t, s', a_{t-1}) \mathbb{P}[s_{t-1}|h_{t-1}]} && \text{Markov}
\end{aligned}$$

- $\mathbb{P}[s_{t-1}|h_{t-1}]$ **with** a_t, o_t **is a sufficient statistic for** s_t
- $b_t \triangleq \mathbb{P}[s_{t-1}|h_{t-1}]$**: belief state at time** t
- b_t **computed recursively using the equation above**

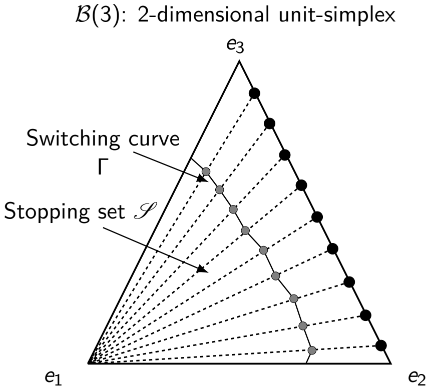
Background: POMDPs





- $b \in \mathcal{B}$, $\mathcal{B}$ is the unit $(|\mathcal{S}| - 1)$-simplex
- **To characterize π^*, partition $\mathcal{B}$ based on $\pi^*(a|b)$**
  - e.g. stopping set $\mathscr{S}$ and continuation set $\mathscr{C}$

Background: POMDPs



- $b \in \mathcal{B}$, $\mathcal{B}$ is the unit $(|\mathcal{S}| - 1)$-simplex
- **To characterize π^*, partition $\mathcal{B}$ based on $\pi^*(a|b)$**
  - e.g. stopping set $\mathscr{S}$ and continuation set $\mathscr{C}$

Background: POMDPs

- $|\mathcal{B}| = \infty$, high-dimensional ($|\mathcal{S}|$) continuous vector
- Infinite set of deterministic policies: $\max_{\pi:\mathcal{B}\to\mathcal{A}} \mathbb{E}_\pi\left[\sum_t r_t\right]$
- However, only finite set of belief points $b \in \mathcal{B}$ are "reachable".
- Finite horizon $\implies$ finite set of "conditional plans" $\mathcal{H} \to \mathcal{A}$

Background: POMDPs

- $|\mathcal{B}| = \infty$, high-dimensional ($|\mathcal{S}|$) continuous vector
- Infinite set of deterministic policies: $\max_{\pi:\mathcal{B}\to\mathcal{A}} \mathbb{E}_\pi\big[\sum_t r_t\big]$

- However, only finite set of belief points $b \in \mathcal{B}$ are "reachable".
- Finite horizon $\implies$ finite set of "conditional plans" $\mathcal{H} \to \mathcal{A}$

Background: POMDPs

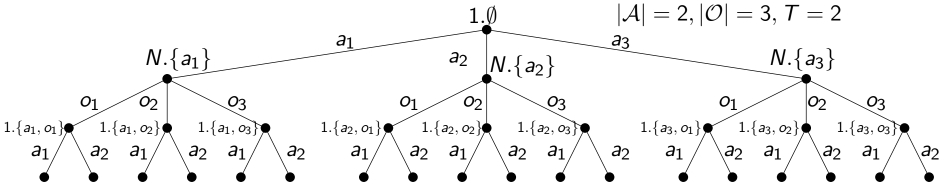
- $|\mathcal{B}| = \infty$, high-dimensional ($|\mathcal{S}|$) continuous vector
- Infinite set of deterministic policies: $\max_{\pi:\mathcal{B}\to\mathcal{A}} \mathbb{E}_\pi \left[\sum_t r_t \right]$

- However, only finite set of belief points $b \in \mathcal{B}$ are "reachable".
- Finite horizon $\implies$ finite set of "conditional plans" $\mathcal{H} \to \mathcal{A}$

Background: POMDPs

- $|\mathcal{B}| = \infty$, high-dimensional ($|\mathcal{S}|$) continuous vector
- Infinite set of deterministic policies: $\max_{\pi:\mathcal{B}\to\mathcal{A}} \mathbb{E}_\pi\left[\sum_t r_t\right]$

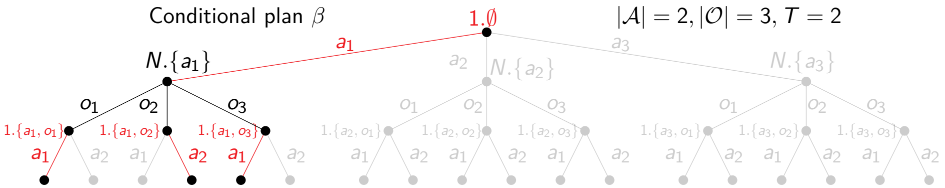
- However, only finite set of belief points $b \in \mathcal{B}$ are "reachable".
- Finite horizon $\implies$ finite set of "conditional plans" $\mathcal{H} \to \mathcal{A}$
  - **Set of pure strategies in an extensive game against nature**



Background: POMDPs

- $|\mathcal{B}| = \infty$, high-dimensional ($|\mathcal{S}|$) continuous vector
- Infinite set of deterministic policies: $\max_{\pi:\mathcal{B}\to\mathcal{A}} \mathbb{E}_\pi\left[\sum_t r_t\right]$

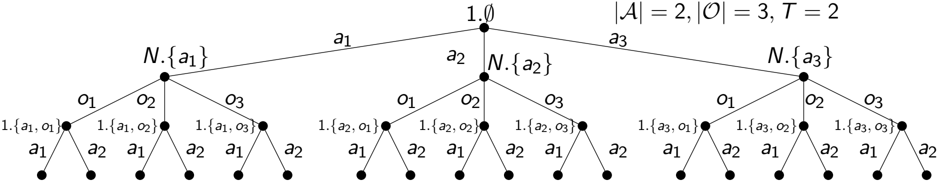
- However, only finite set of belief points $b \in \mathcal{B}$ are "reachable".
- Finite horizon $\implies$ finite set of "conditional plans" $\mathcal{H} \to \mathcal{A}$
  - **Set of pure strategies in an extensive game against nature**



Background: POMDPs

- $|\mathcal{B}| = \infty$, high-dimensional ($|\mathcal{S}|$) continuous vector
- Infinite set of deterministic policies: $\max_{\pi:\mathcal{B}\to\mathcal{A}} \mathbb{E}_\pi\left[\sum_t r_t\right]$

- However, only finite set of belief points $b \in \mathcal{B}$ are "reachable".
- Finite horizon $\implies$ finite set of "conditional plans" $\mathcal{H} \to \mathcal{A}$
  - **Set of pure strategies in an extensive game against nature**



Background: POMDPs

- For each conditional plan $\beta \in \Gamma$:
  - Define vector $\alpha^\beta \in \mathbb{R}^{|\mathcal{S}|}$ such that $\alpha_i^\beta = V^\beta(i)$
  - $\implies V^\beta(b) = b^T \alpha^\beta$ (linear in b).
- Thus, $V^*(b) = \max_{\beta \in \Gamma} b^T \alpha^\beta$ (piece-wise linear and convex[1])

[1] Edward J. Sondik. "The Optimal Control of Partially Observable Markov Processes Over the Infinite Horizon: Discounted Costs". In: *Operations Research* 26.2 (1978), pp. 282–304. ISSN: 0030364X, 15265463. URL: http://www.jstor.org/stable/169635.

Background: POMDPs

- For each conditional plan $\beta \in \Gamma$:
  - Define vector $\alpha^\beta \in \mathbb{R}^{|\mathcal{S}|}$ such that $\alpha_i^\beta = V^\beta(i)$
  - $\implies V^\beta(b) = b^T \alpha^\beta$ (linear in b).
- Thus, $V^*(b) = \max_{\beta \in \Gamma} b^T \alpha^\beta$ (piece-wise linear and convex[2])

[2] Edward J. Sondik. "The Optimal Control of Partially Observable Markov Processes Over the Infinite Horizon: Discounted Costs". In: *Operations Research* 26.2 (1978), pp. 282–304. ISSN: 0030364X, 15265463. URL: http://www.jstor.org/stable/169635.

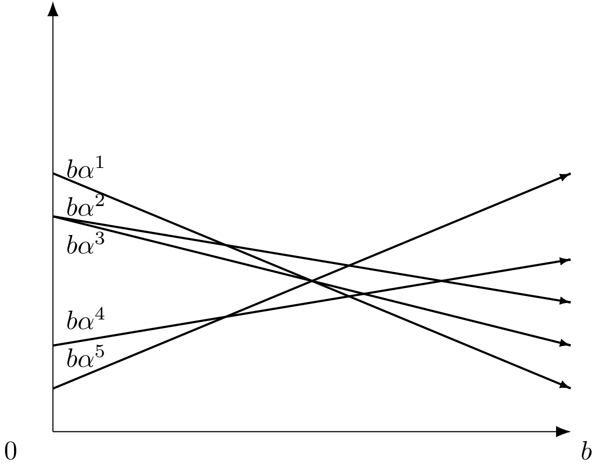
Background: POMDPs

- For each conditional plan $\beta \in \Gamma$:
  - Define vector $\alpha^\beta \in \mathbb{R}^{|\mathcal{S}|}$ such that $\alpha_i^\beta = V^\beta(i)$
  - $\implies V^\beta(b) = b^T \alpha^\beta$ (linear in b).
- Thus, $V^*(b) = \max_{\beta \in \Gamma} b^T \alpha^\beta$ (piece-wise linear and convex[3])

[3] Edward J. Sondik. "The Optimal Control of Partially Observable Markov Processes Over the Infinite Horizon: Discounted Costs". In: *Operations Research* 26.2 (1978), pp. 282–304. ISSN: 0030364X, 15265463. URL: http://www.jstor.org/stable/169635.

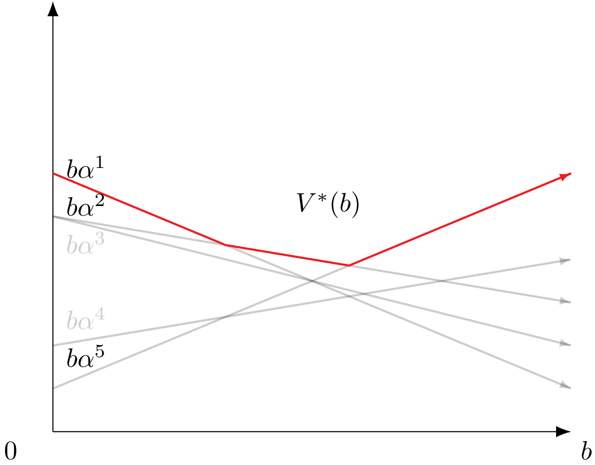
Background: POMDPs

- For each conditional plan $\beta \in \Gamma$:
  - Define vector $\alpha^\beta \in \mathbb{R}^{|\mathcal{S}|}$ such that $\alpha^\beta_i = V^\beta(i)$
  - $\implies V^\beta(b) = b^T \alpha^\beta$ (linear in b).
- Thus, $V^*(b) = \max_{\beta \in \Gamma} b^T \alpha^\beta$ (piece-wise linear and convex[4])



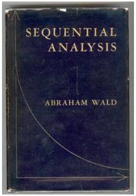
Background: POMDPs

- For each conditional plan $\beta \in \Gamma$:
  - Define vector $\alpha^\beta \in \mathbb{R}^{|\mathcal{S}|}$ such that $\alpha_i^\beta = V^\beta(i)$
  - $\implies V^\beta(b) = b^T \alpha^\beta$ (linear in b).
- Thus, $V^*(b) = \max_{\beta \in \Gamma} b^T \alpha^\beta$ (piece-wise linear and convex[5])



Background: Optimal Stopping

- **History**:
  - Studied in the 18th century to analyze a gambler's fortune
  - Formalized by Abraham Wald in 1947[6]
  - Since then it has been generalized and developed by (Chow[7], Shiryaev & Kolmogorov[8], Bather[9], Bertsekas[10], etc.)



[6] Abraham Wald. *Sequential Analysis*. Wiley and Sons, New York, 1947.

[7] Y. Chow, H. Robbins, and D. Siegmund. "Great expectations: The theory of optimal stopping". In: 1971.

[8] Albert N. Shirayev. *Optimal Stopping Rules*. Reprint of russian edition from 1969. Springer-Verlag Berlin, 2007.

[9] John Bather. *Decision Theory: An Introduction to Dynamic Programming and Sequential Decisions*. USA: John Wiley and Sons, Inc., 2000. ISBN: 0471976490.

[10] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. 3rd. Vol. I. Belmont, MA, USA: Athena Scientific, 2005.

Background: Optimal Stopping

- **The General Problem**:
  - A stochastic process $(s_t)_{t=1}^{T}$ is observed sequentially
  - Two options per t: (i) continue to observe; or (ii) stop
  - Find the *optimal stopping time* τ^*:

$$\tau^* = \arg\max_{\tau} \mathbb{E}_{\tau}\left[\sum_{t=1}^{\tau-1} \gamma^{t-1}\mathcal{R}^{C}_{s_t s_{t+1}} + \gamma^{\tau-1}\mathcal{R}^{S}_{s_\tau s_\tau}\right] \tag{1}$$

  where $\mathcal{R}^{S}_{ss'}$ & $\mathcal{R}^{C}_{ss'}$ are the stop/continue rewards

- **Solution approaches:** the *Markovian approach* and the *martingale approach*.

Background: Optimal Stopping

- **The General Problem**:
  - A stochastic process $(s_t)_{t=1}^{T}$ is observed sequentially
  - Two options per t: (i) continue to observe; or (ii) stop
  - Find the *optimal stopping time* τ^*:

$$\tau^* = \arg\max_{\tau} \mathbb{E}_\tau \left[\sum_{t=1}^{\tau-1} \gamma^{t-1} \mathcal{R}^{C}_{s_t s_{t+1}} + \gamma^{\tau-1} \mathcal{R}^{S}_{s_\tau s_\tau} \right] \qquad (2)$$

  where $\mathcal{R}^{S}_{ss'}$ & $\mathcal{R}^{C}_{ss'}$ are the stop/continue rewards

- **Solution approaches:** the *Markovian approach* and the *martingale approach*.

Background: Optimal Stopping

- **The Markovian approach**:
  - Model the problem as a MDP or POMDP
  - A policy π^* that satisfies the Bellman-Wald equation is optimal:

$$\pi^*(s) = \underset{\{S,C\}}{\arg\max} \left[\underbrace{\mathbb{E}\left[\mathcal{R}_s^S\right]}_{\text{stop}}, \underbrace{\mathbb{E}\left[\mathcal{R}_s^C + \gamma V^*(s')\right]}_{\text{continue}} \right] \quad \forall s \in \mathcal{S}$$

  - Solve by backward induction, dynamic programming, or reinforcement learning

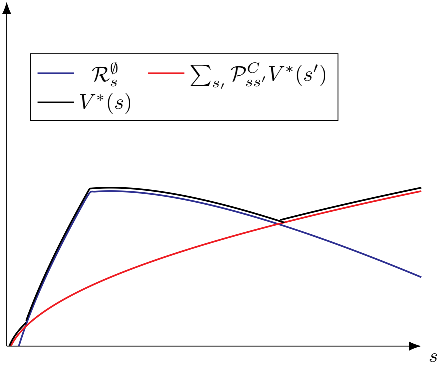
- **The martingale approach**:
  - Model the state process as an arbitrary stochastic process
  - The reward of the optimal stopping time is given by the *smallest supermartingale that stochastically dominates the process*, called the Snell envelope [13].

Background: Optimal Stopping

- **The Markovian approach**:
  - Assume all rewards are received upon stopping: $R_s^{\emptyset}$
  - $V^*(s)$ **majorizes** $R_s^{\emptyset}$ if $V^*(s) \geq R_s^{\emptyset}\ \forall s \in \mathcal{S}$
  - $V^*(s)$ is **excessive** if $V^*(s) \geq \sum_{s'} \mathcal{P}^{C}_{s's} V^*(s')\ \forall s \in \mathcal{S}$
  - Theorem: $V^*(s)$ is the minimal excessive function which majorizes $R_s^{\emptyset}$.
- **The martingale approach**:
  - Model the state process as an arbitrary stochastic process
  - The reward of the optimal stopping time is given by the *smallest supermartingale that stochastically dominates the process*, called the Snell envelope [13].

Background: Optimal Stopping

- **The Markovian approach**:
  - Assume all rewards are received upon stopping: $R_s^\emptyset$
  - $V^*(s)$ **majorizes** $R_s^\emptyset$ if $V^*(s) \geq R_s^\emptyset \ \forall s \in \mathcal{S}$
  - $V^*(s)$ is **excessive** if $V^*(s) \geq \sum_{s'} \mathcal{P}^C_{s's} V^*(s') \ \forall s \in \mathcal{S}$
  - $V^*(s)$ is the minimal excessive function which majorizes $R_s^\emptyset$.



Background: Optimal Stopping

- **The Markovian approach**:
  - Assume all rewards are received upon stopping: $R_s^{\emptyset}$
  - $V^*(s)$ **majorizes** $R_s^{\emptyset}$ if $V^*(s) \geq R_s^{\emptyset}\ \forall s \in \mathcal{S}$
  - $V^*(s)$ is **excessive** if $V^*(s) \geq \sum_{s'} \mathcal{P}_{s's}^{C} V^*(s')\ \forall s \in \mathcal{S}$
  - $V^*(s)$ is the minimal excessive function which majorizes $R_s^{\emptyset}$.

- **The martingale approach**:
  - Model the state process as an arbitrary stochastic process

  - The reward of the optimal stopping time is given by the *smallest supermartingale that stochastically dominates the process*, called the Snell envelope[11].

[11] J. L. Snell. "Applications of martingale system theorems". In: *Transactions of the American Mathematical Society* 73 (1952), pp. 293–312.

Background: Optimal Stopping

- **Applications & Use Cases**:
  - Hypothesis testing[12]
  - Change detection[13],
  - Selling decisions[14],
  - Queue management[15],
  - Industrial control[16],
  - Advertisement scheduling[17], etc.

---

[12] Abraham Wald. *Sequential Analysis*. Wiley and Sons, New York, 1947.

[13] Alexander G. Tartakovsky et al. "Detection of intrusions in information systems by sequential change-point methods". In: *Statistical Methodology* 3.3 (2006). ISSN: 1572-3127. DOI: https://doi.org/10.1016/j.stamet.2005.05.003. URL: https://www.sciencedirect.com/science/article/pii/S1572312705000493.

[14] Jacques du Toit and Goran Peskir. "Selling a stock at the ultimate maximum". In: *The Annals of Applied Probability* 19.3 (2009). ISSN: 1050-5164. DOI: 10.1214/08-aap566. URL: http://dx.doi.org/10.1214/08-AAP566.

[15] Arghyadip Roy et al. "Online Reinforcement Learning of Optimal Threshold Policies for Markov Decision Processes". In: *CoRR* (2019). http://arxiv.org/abs/1912.10325. eprint: 1912.10325.

[16] Maben Rabi and Karl H. Johansson. "Event-Triggered Strategies for Industrial Control over Wireless Networks". In: *Proceedings of the 4th Annual International Conference on Wireless Internet*. WICON '08. Maui, Hawaii, USA: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008. ISBN: 9789639799363.

[17] Vikram Krishnamurthy, Anup Aprem, and Sujay Bhatt. "Multiple stopping time POMDPs: Structural results & application in interactive advertising on social media". In: *Automatica* 95 (2018), pp. 385–398. ISSN: 0005-1098. DOI: https://doi.org/10.1016/j.automatica.2018.06.013. URL: https://www.sciencedirect.com/science/article/pii/S0005109818303054.

Background: Optimal Stopping

- **Applications & Use Cases**:
  - Hypothesis testing[18]
  - Change detection[19],
  - Selling decisions[20],
  - Queue management[21],
  - Industrial control[22],
  - Advertisement scheduling,
  - **Intrusion prevention**[23] etc.

---

[18] Abraham Wald. *Sequential Analysis*. Wiley and Sons, New York, 1947.

[19] Alexander G. Tartakovsky et al. "Detection of intrusions in information systems by sequential change-point methods". In: *Statistical Methodology* 3.3 (2006). ISSN: 1572-3127. DOI: https://doi.org/10.1016/j.stamet.2005.05.003. URL: https://www.sciencedirect.com/science/article/pii/S1572312705000493.

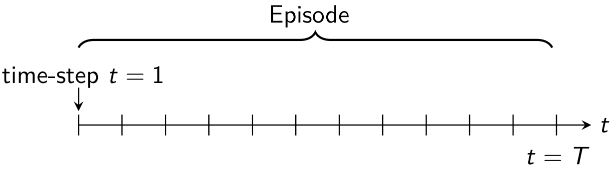
[20] Jacques du Toit and Goran Peskir. "Selling a stock at the ultimate maximum". In: *The Annals of Applied Probability* 19.3 (2009). ISSN: 1050-5164. DOI: 10.1214/08-aap566. URL: http://dx.doi.org/10.1214/08-AAP566.

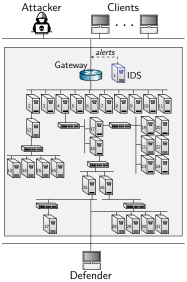
[21] Arghyadip Roy et al. "Online Reinforcement Learning of Optimal Threshold Policies for Markov Decision Processes". In: *CoRR* (2019). http://arxiv.org/abs/1912.10325. eprint: 1912.10325.

[22] Maben Rabi and Karl H. Johansson. "Event-Triggered Strategies for Industrial Control over Wireless Networks". In: *Proceedings of the 4th Annual International Conference on Wireless Internet*. WICON '08. Maui, Hawaii, USA: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008. ISBN: 9789639799363.

[23] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

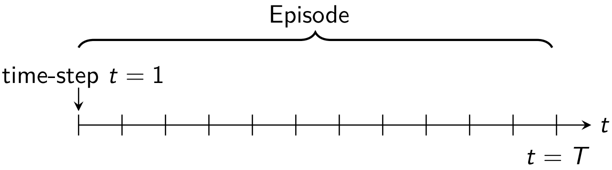
Formulating Intrusion Prevention as a Stopping Problem

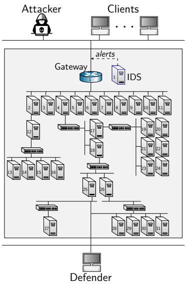




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations, when is it optimal to stop?**
- We formalize this problem with a POMDP

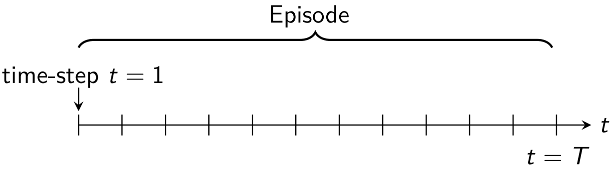
Formulating Intrusion Prevention as a Stopping Problem

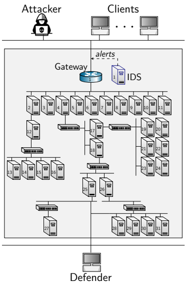




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations, when is it optimal to stop?**
- We formalize this problem with a POMDP

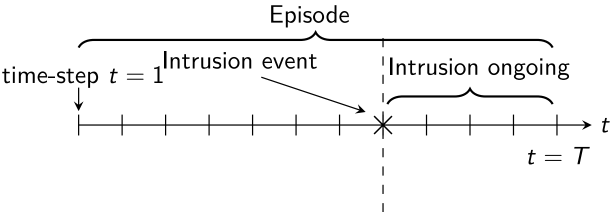
Formulating Intrusion Prevention as a Stopping Problem

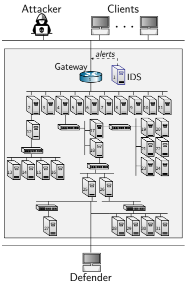




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations, when is it optimal to stop?**
- We formalize this problem with a POMDP

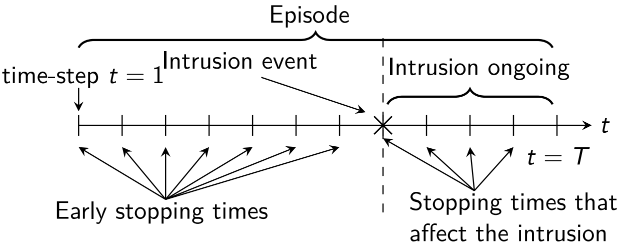
Formulating Intrusion Prevention as a Stopping Problem

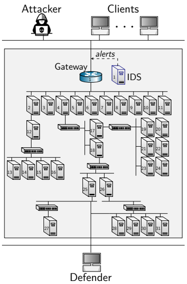




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations, when is it optimal to stop?**
- We formalize this problem with a POMDP

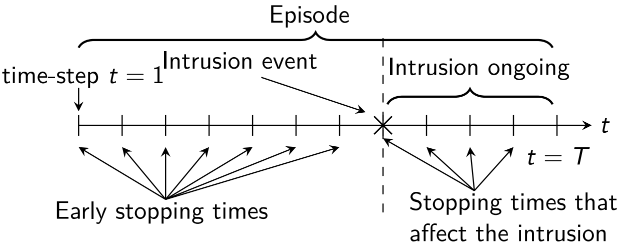
Formulating Intrusion Prevention as a Stopping Problem

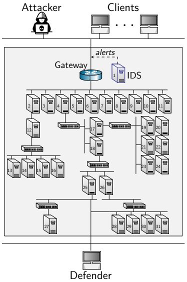




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations, when is it optimal to stop?**
- We formalize this problem with a POMDP

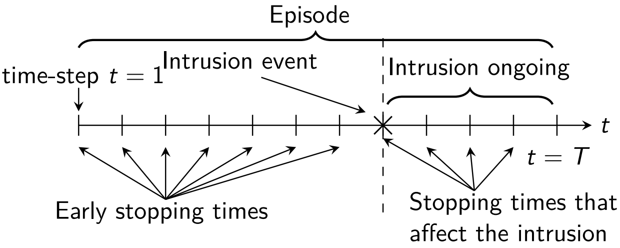
Formulating Intrusion Prevention as a Stopping Problem

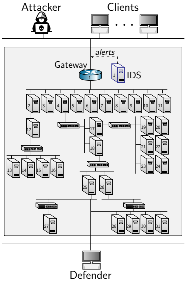




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations, when is it optimal to stop?**
- We formalize this problem with a POMDP

Formulating Intrusion Prevention as a Stopping Problem

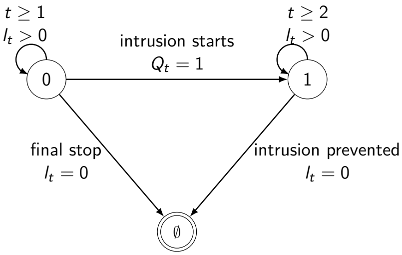




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations, when is it optimal to stop?**
- We formalize this problem with a POMDP

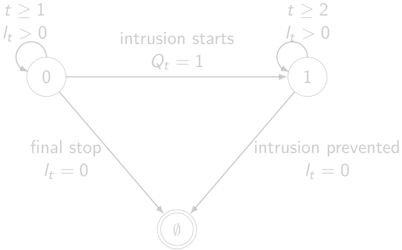
A Partially Observed MDP Model for the Defender

- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1, .., L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$



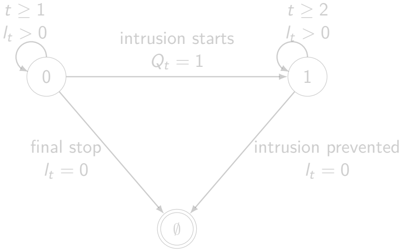
A Partially Observed MDP Model for the Defender

- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1, .., L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta} \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t) \right]$, $T_\emptyset$



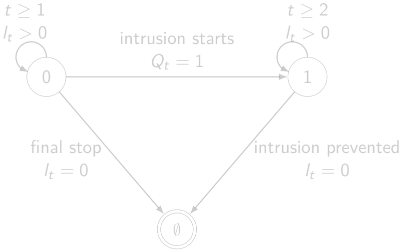
A Partially Observed MDP Model for the Defender

- **States:**
  - Intrusion state $s_t \in \{0,1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1,..,L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$



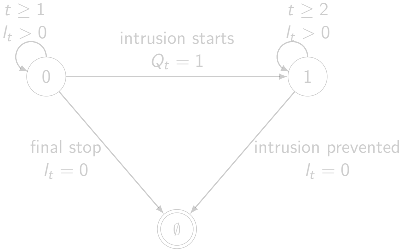
A Partially Observed MDP Model for the Defender

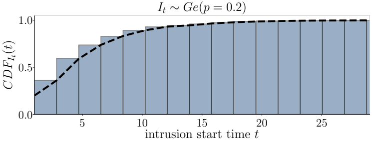
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1, .., L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$



A Partially Observed MDP Model for the Defender

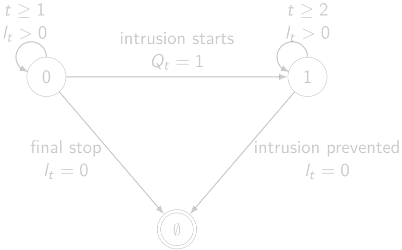
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1, .., L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta} \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t) \right]$, $T_\emptyset$





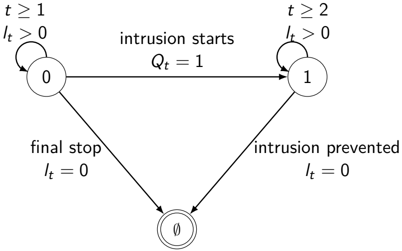
A Partially Observed MDP Model for the Defender

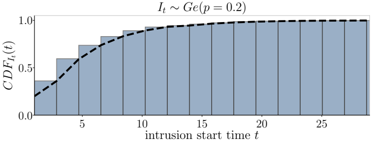
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1, .., L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta} \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t) \right]$, $T_\emptyset$



A Partially Observed MDP Model for the Defender

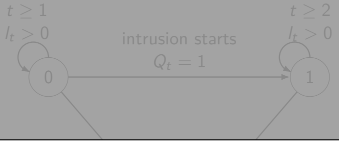
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1, .., L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta} \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$





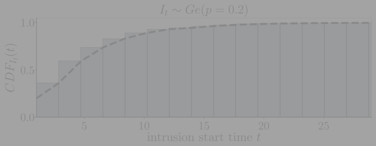
A Partially Observed MDP Model for the Defender

- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - Severe/Warning IDS Alerts $(\Delta x, \Delta y)$, Login attempts Δz, stops remaining $l_t \in \{1, .., L\}$, $f_{XYZ}(\Delta x, \Delta y, \Delta z | s_t)$



We analyze the structure of π^* using POMDP & stopping theory

  - Reward: security and service. Penalty: false alarms and intrusions
- **Transition probabilities:**
  - Bernoulli process $(Q_t)_{t=1}^{T} \sim Ber(p)$ defines intrusion start $I_t \sim Ge(p)$
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi_\theta} \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t) \right]$, $T_\emptyset$



Outline

- **Use Case & Approach:**
  - Intrusion Prevention
  - System identification
  - Reinforcement learning and dynamic programming
- **Formal Model & Background:**
  - Background: POMDPs and optimal stopping
  - Multiple Stopping Problem POMDP
- **Structure of π^***
  - Structural result: Multi-Threshold policy
  - Stopping sets $\mathscr{S}_l$ are connected and nested
  - Conditions for Bayesian filter to be monotone in b
  - Existence of optimal multi-threshold policy π^*_l
- **Conclusion**
  - Numerical evaluation results
  - Conclusion & Future work

Structural Result: Optimal Multi-Threshold Policy

Theorem

Given the intrusion prevention POMDP, the following holds:

1. $\mathscr{S}_{l-1} \subseteq \mathscr{S}_l$ *for* $l = 2, \ldots L$.
2. *If* $L = 1$*, there exists an optimal threshold* $\alpha^* \in [0,1]$ *and an optimal policy of the form:*

$$\pi_L^*(b(1)) = S \iff b(1) \geq \alpha^* \tag{3}$$

3. *If* $L \geq 1$ *and* f_{XYZ} *is totally positive of order 2 (TP2), there exists* L *optimal thresholds* $\alpha_l^* \in [0,1]$ *and an optimal policy of the form:*

$$\pi_l^*(b(1)) = S \iff b(1) \geq \alpha_l^*, \qquad l = 1, \ldots, L \tag{4}$$

where α_l^* *is decreasing in* l.

Structural Result: Optimal Multi-Threshold Policy

Theorem

Given the intrusion prevention POMDP, the following holds:

1. $\mathscr{S}_{l-1} \subseteq \mathscr{S}_l$ *for* $l = 2, \ldots L$.
2. *If* $L = 1$, *there exists an optimal threshold* $\alpha^* \in [0, 1]$ *and an optimal policy of the form:*

$$\pi^*_L(b(1)) = S \iff b(1) \geq \alpha^* \tag{5}$$

3. *If* $L \geq 1$ *and* f_{XYZ} *is totally positive of order 2 (TP2), there exists* L *optimal thresholds* $\alpha^*_l \in [0, 1]$ *and an optimal policy of the form:*

$$\pi^*_l(b(1)) = S \iff b(1) \geq \alpha^*_l, \qquad l = 1, \ldots, L \tag{6}$$

where α^*_l *is decreasing in* l.

Structural Result: Optimal Multi-Threshold Policy

Theorem

Given the intrusion prevention POMDP, the following holds:

1. $\mathscr{S}_{l-1} \subseteq \mathscr{S}_l$ *for* $l = 2, \ldots L$.
2. *If* $L = 1$, *there exists an optimal threshold* $\alpha^* \in [0, 1]$ *and an optimal policy of the form:*

$$\pi_L^*(b(1)) = S \iff b(1) \geq \alpha^* \tag{7}$$

3. *If* $L \geq 1$ *and* f_{XYZ} *is totally positive of order 2 (TP2), there exists* L *optimal thresholds* $\alpha_l^* \in [0, 1]$ *and an optimal policy of the form:*

$$\pi_l^*(b(1)) = S \iff b(1) \geq \alpha_l^*, \qquad l = 1, \ldots, L \tag{8}$$

where α_l^* *is decreasing in* l.

Structural Result: Optimal Multi-Threshold Policy

Theorem

Given the intrusion prevention POMDP, the following holds:

1. $\mathscr{S}_{l-1} \subseteq \mathscr{S}_l$ *for* $l = 2, \ldots L$.
2. *If* $L = 1$, *there exists an optimal threshold* $\alpha^* \in [0, 1]$ *and an optimal policy of the form:*

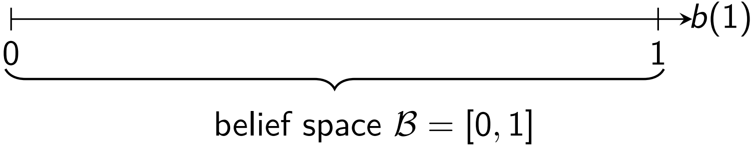
$$\pi_L^*(b(1)) = S \iff b(1) \geq \alpha^* \tag{9}$$

3. *If* $L \geq 1$ *and* f_{XYZ} *is totally positive of order 2 (TP2), there exists* L *optimal thresholds* $\alpha_l^* \in [0, 1]$ *and an optimal policy of the form:*

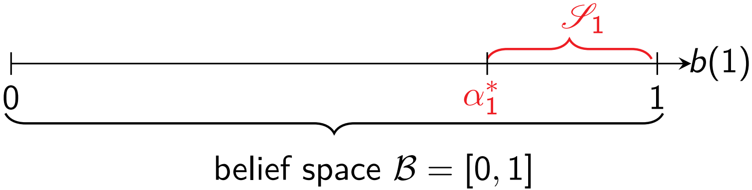
$$\pi_l^*(b(1)) = S \iff b(1) \geq \alpha_l^*, \qquad l = 1, \ldots, L \tag{10}$$

where α_l^* *is decreasing in* l.

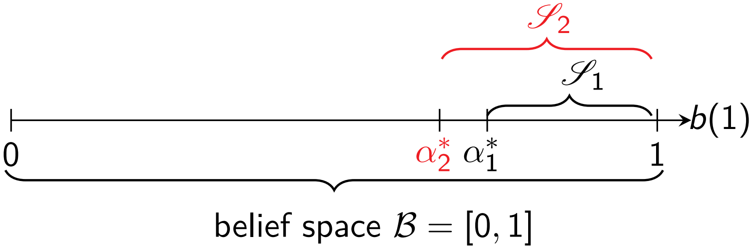
Structural Result: Optimal Multi-Threshold Policy



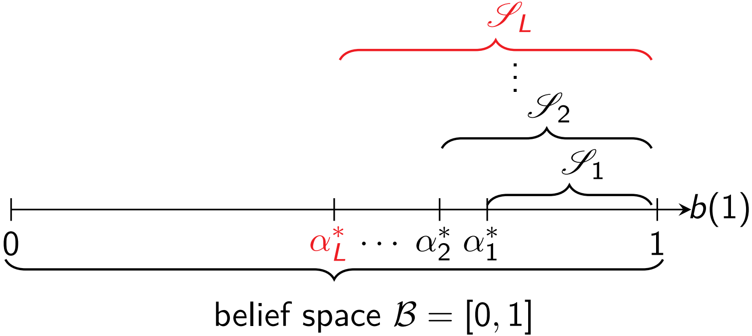
Structural Result: Optimal Multi-Threshold Policy



Structural Result: Optimal Multi-Threshold Policy



Structural Result: Optimal Multi-Threshold Policy



Proofs: $\mathscr{S}_1$ is convex[24]

- $\mathscr{S}_1$ is convex if:
  - for any two belief states $b_1, b_2 \in \mathscr{S}_1$
  - any convex combination of b_1, b_2 is also in $\mathscr{S}_1$
  - i.e. $b_1, b_2 \in \mathscr{S}_1 \implies \lambda b_1 + (1-\lambda)b_2 \in \mathscr{S}_1$ for $\lambda \in [0,1]$.

- Since $V^*(b)$ is convex:

$$V^*(\lambda b_1 + (1-\lambda)b_2) \leq \lambda V^*(b_1) + (1-\lambda)V(b_2)$$

- Since $b_1, b_2 \in \mathscr{S}_1$:

$$V^*(b_1) = Q^*(b_1, S) \qquad S{=}\text{stop}$$
$$V^*(b_2) = Q^*(b_2, S) \qquad S{=}\text{stop}$$

[24] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[25]

- $\mathscr{S}_1$ is convex if:
  - for any two belief states $b_1, b_2 \in \mathscr{S}_1$
  - any convex combination of b_1, b_2 is also in $\mathscr{S}_1$
  - i.e. $b_1, b_2 \in \mathscr{S}_1 \implies \lambda b_1 + (1-\lambda)b_2 \in \mathscr{S}_1$ for $\lambda \in [0,1]$.

- Since $V^*(b)$ is convex:

$$V^*(\lambda b_1 + (1-\lambda)b_2) \leq \lambda V^*(b_1) + (1-\lambda)V(b_2)$$

- Since $b_1, b_2 \in \mathscr{S}_1$:

$$\begin{aligned} V^*(b_1) &= Q^*(b_1, S) && S\text{=stop} \\ V^*(b_2) &= Q^*(b_2, S) && S\text{=stop} \end{aligned}$$

[25] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[26]

- $\mathscr{S}_1$ is convex if:
  - for any two belief states $b_1, b_2 \in \mathscr{S}_1$
  - any convex combination of b_1, b_2 is also in $\mathscr{S}_1$
  - i.e. $b_1, b_2 \in \mathscr{S}_1 \implies \lambda b_1 + (1-\lambda) b_2 \in \mathscr{S}_1$ for $\lambda \in [0,1]$.

- Since $V^*(b)$ is convex:

$$V^*(\lambda b_1 + (1-\lambda) b_2) \leq \lambda V^*(b_1) + (1-\lambda) V(b_2)$$

- Since $b_1, b_2 \in \mathscr{S}_1$:

$$\begin{aligned} V^*(b_1) &= Q^*(b_1, S) && S\text{=stop} \\ V^*(b_2) &= Q^*(b_2, S) && S\text{=stop} \end{aligned}$$

[26] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[27]

Proof.

Assume $b_1, b_2 \in \mathscr{S}_1$. Then for any $\lambda \in [0,1]$:

$$
\begin{aligned}
V^*(\lambda b_1(1) + (1-\lambda) b_2(1)) &\leq \lambda V^*(b_1(1)) + (1-\lambda) V^*(b_2(1)) \\
&= \lambda Q^*(b_1, S) + (1-\lambda) Q^*(b_2, S)
\end{aligned}
$$

□

[27] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[28]

Proof.

Assume $b_1, b_2 \in \mathscr{S}_1$. Then for any $\lambda \in [0,1]$:

$$\begin{aligned}
V^*(\lambda b_1(1) + (1-\lambda) b_2(1)) &\leq \lambda V^*(b_1(1)) + (1-\lambda) V^*(b_2(1)) \\
&= \lambda Q^*(b_1, S) + (1-\lambda) Q^*(b_2, S) \\
&= \lambda \mathcal{R}^{\emptyset}_{b_1} + (1-\lambda)\mathcal{R}^{\emptyset}_{b_2} \\
&= \sum_s (\lambda b_1(s) + (1-\lambda) b_2(s)) \mathcal{R}^{\emptyset}_s
\end{aligned}$$

□

[28] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[29]

Proof.

Assume $b_1, b_2 \in \mathscr{S}_1$. Then for any $\lambda \in [0,1]$:

$$
\begin{aligned}
V^*(\lambda b_1(1) + (1-\lambda) b_2(1)) &\leq \lambda V^*(b_1(1)) + (1-\lambda) V^*(b_2(1)) \\
&= \lambda Q^*(b_1, S) + (1-\lambda) Q^*(b_2, S) \\
&= \lambda \mathcal{R}^{\emptyset}_{b_1} + (1-\lambda)\mathcal{R}^{\emptyset}_{b_2} \\
&= \sum_s (\lambda b_1(s) + (1-\lambda) b_2(s)) \mathcal{R}^{\emptyset}_s \\
&= Q^*(\lambda b_1 + (1-\lambda) b_2, S)
\end{aligned}
$$

□

[29] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[30]

Proof.

Assume $b_1, b_2 \in \mathscr{S}_1$. Then for any $\lambda \in [0,1]$:

$$\begin{aligned}
V^*(\lambda b_1(1) + (1-\lambda) b_2(1)) &\leq \lambda V^*(b_1(1)) + (1-\lambda) V^*(b_2(1)) \\
&= \lambda Q^*(b_1, S) + (1-\lambda) Q^*(b_2, S) \\
&= \lambda \mathcal{R}^{\emptyset}_{b_1} + (1-\lambda)\mathcal{R}^{\emptyset}_{b_2} \\
&= \sum_s (\lambda b_1(s) + (1-\lambda) b_2(s)) \mathcal{R}^{\emptyset}_s \\
&= Q^*(\lambda b_1 + (1-\lambda) b_2, S) \\
&\leq V^*(\lambda b_1(1) + (1-\lambda) b_2(1))
\end{aligned}$$

the last inequality is because V^* is optimal. The second-to-last is because there is just a single stop. □

[30] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[31]

Proof.

Assume $b_1, b_2 \in \mathscr{S}_1$. Then for any $\lambda \in [0,1]$:

$$\begin{aligned} V^*(\lambda b_1(1) + (1-\lambda) b_2(1)) &\leq \lambda V^*(b_1(1)) + (1-\lambda) V^*(b_2(1)) \\ &= \lambda Q^*(b_1, S) + (1-\lambda) Q^*(b_2, S) \\ &= Q^*(\lambda b_1 + (1-\lambda) b_2, S) \\ &\leq V^*(\lambda b_1(1) + (1-\lambda) b_2(1)) \end{aligned}$$

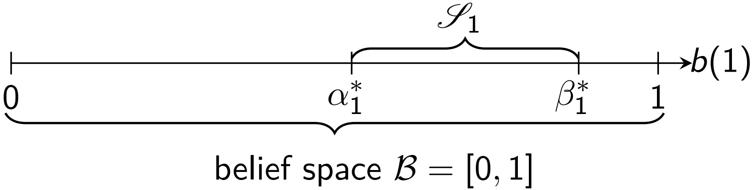
the last inequality is because V^* is optimal. The second-to-last is because there is just a single stop. Hence:

$$Q^*(\lambda b_1 + (1-\lambda) b_2, S) = V^*(\lambda b_1(1) + (1-\lambda) b_2(1))$$

$b_1, b_2 \in \mathscr{S}_1 \implies (\lambda b_1 + (1-\lambda)) \in \mathscr{S}_1$. Therefore $\mathscr{S}_1$ is convex. $\square$

[31] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: $\mathscr{S}_1$ is convex[32]



[32] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: Single-threshold policy is optimal if $L = 1$[33]

- In our case, $\mathcal{B} = [0,1]$. We know $\mathscr{S}_1$ is a convex subset of $\mathcal{B}$.
- Consequence, $\mathscr{S}_1 = [\alpha^*, \beta^*]$. We show that $\beta^* = 1$.
- If $b(1) = 1$, using our definition of the reward function, the Bellman equation states:

$$\pi^*(1) \in \arg\max_{\{S,C\}} \Bigg[\underbrace{150 + V^*(\emptyset)}_{a=S}, \underbrace{-90 + \sum_{o \in \mathcal{O}} \mathcal{Z}(o,1,C) V^*(b_C^o(1))}_{a=C} \Bigg]$$

$$= \arg\max_{\{S,C\}} \Bigg[\underbrace{150}_{a=S}, \underbrace{-90 + V^*(1)}_{a=C} \Bigg] = S \quad \text{i.e } \pi^*(1) = \text{Stop}$$

- Hence $1 \in \mathscr{S}_1$. It follows that $\mathscr{S}_1 = [\alpha^*, 1]$ and:

$$\pi^*(b(1)) = \begin{cases} S & \text{if } b(1) \geq \alpha^* \\ C & \text{otherwise} \end{cases}$$

[33] Kim Hammar and Rolf Stadler. "Learning Intrusion Prevention Policies through Optimal Stopping". In: *International Conference on Network and Service Management (CNSM 2021)*. https://arxiv.org/pdf/2106.07160.pdf. Izmir, Turkey, 2021.

Proofs: Single-threshold policy is optimal if $L = 1$[34]

- In our case, $\mathcal{B} = [0, 1]$. We know $\mathscr{S}_1$ is a convex subset of $\mathcal{B}$.
- Consequence, $\mathscr{S}_1 = [\alpha^*, \beta^*]$. We show that $\beta^* = 1$.
- If $b(1) = 1$, using our definition of the reward function, the Bellman equation states:

$$\pi^*(1) \in \arg\max_{\{S,C\}} \Bigg[\underbrace{150 + V^*(\emptyset)}_{a=S}, \underbrace{-90 + \sum_{o \in \mathcal{O}} \mathcal{Z}(o, 1, C) V^*(b_C^o(1))}_{a=C}\Bigg]$$

$$= \arg\max_{\{S,C\}} \Bigg[\underbrace{150}_{a=S}, \underbrace{-90 + V^*(1)}_{a=C}\Bigg] = S \quad \text{i.e } \pi^*(1) = \text{Stop}$$

- Hence $1 \in \mathscr{S}_1$. It follows that $\mathscr{S}_1 = [\alpha^*, 1]$ and:

$$\pi^*(b(1)) = \begin{cases} S & \text{if } b(1) \geq \alpha^* \\ C & \text{otherwise} \end{cases}$$

[34] Kim Hammar and Rolf Stadler. "Learning Intrusion Prevention Policies through Optimal Stopping". In: *International Conference on Network and Service Management (CNSM 2021)*. https://arxiv.org/pdf/2106.07160.pdf. Izmir, Turkey, 2021.

Proofs: Single-threshold policy is optimal if $L = 1$[35]

- In our case, $\mathcal{B} = [0,1]$. We know $\mathscr{S}_1$ is a convex subset of $\mathcal{B}$.
- Consequence, $\mathscr{S}_1 = [\alpha^*, \beta^*]$. We show that $\beta^* = 1$.
- If $b(1) = 1$, using our definition of the reward function, the Bellman equation states:

$$\pi^*(1) \in \arg\max_{\{S,C\}} \Bigg[\underbrace{150 + V^*(\emptyset)}_{a=S}, \underbrace{-90 + \sum_{o \in \mathcal{O}} \mathcal{Z}(o,1,C) V^*(b_C^o(1))}_{a=C} \Bigg]$$

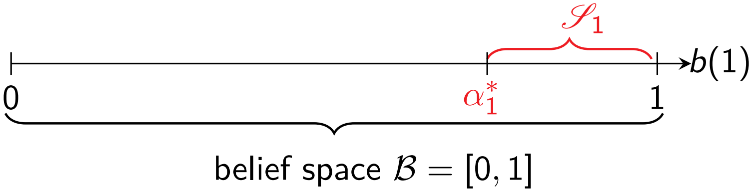
$$= \arg\max_{\{S,C\}} \Bigg[\underbrace{150}_{a=S}, \underbrace{-90 + V^*(1)}_{a=C} \Bigg] = S \quad \text{i.e } \pi^*(1) = \text{Stop}$$

- Hence $1 \in \mathscr{S}_1$. It follows that $\mathscr{S}_1 = [\alpha^*, 1]$ and:

$$\pi^*(b(1)) = \begin{cases} S & \text{if } b(1) \geq \alpha^* \\ C & \text{otherwise} \end{cases}$$

[35] Kim Hammar and Rolf Stadler. "Learning Intrusion Prevention Policies through Optimal Stopping". In: *International Conference on Network and Service Management (CNSM 2021)*. https://arxiv.org/pdf/2106.07160.pdf. Izmir, Turkey, 2021.

Proofs: Single-threshold policy is optimal if $L = 1$



Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[36]

- If $b(1) \in \mathscr{S}_{l-1}$, we use the Bellman eq. to obtain:
- $$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} \geq \sum_o \mathbb{P}^o_{b(1)}\Big(V^*_{l-1}(b^o(1)) - V^*_{l-2}(b^o(1))\Big)$$
- **We show that LHS is non-decreasing in l and RHS is non-increasing in l.**
- LHS is non-decreasing by definition of reward function.
- We show that RHS is non-increasing by induction on $k = 0, 1 \ldots$ where k is the iteration of value iteration.
- We know $\lim_{k\to\infty} V^k(b) = V^*(b)$.
- Define $W^k_l(b(1)) = V^k_l(b(1)) - V^k_{l-1}(b(1))$

[36] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[37]

- If $b(1) \in \mathscr{S}_{l-1}$, we use the Bellman eq. to obtain:
- $$\mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} \geq \sum_{o} \mathbb{P}^{o}_{b(1)}\Big(V^{*}_{l-1}(b^{o}(1)) - V^{*}_{l-2}(b^{o}(1))\Big)$$
- **We show that LHS is non-decreasing in l and RHS is non-increasing in l.**
- LHS is non-decreasing by definition of reward function.
- We show that RHS is non-increasing by induction on $k = 0, 1 \ldots$ where k is the iteration of value iteration.
- We know $\lim_{k\to\infty} V^{k}(b) = V^{*}(b)$.
- Define $W^{k}_{l}(b(1)) = V^{k}_{l}(b(1)) - V^{k}_{l-1}(b(1))$

[37] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[38]

- If $b(1) \in \mathscr{S}_{l-1}$, we use the Bellman eq. to obtain:
- $$\mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} \geq \sum_{o} \mathbb{P}^{o}_{b(1)} \Big(V^{*}_{l-1}(b^{o}(1)) - V^{*}_{l-2}(b^{o}(1)) \Big)$$
- **We show that LHS is non-decreasing in l and RHS is non-increasing in l.**
- LHS is non-decreasing by definition of reward function.
- We show that RHS is non-increasing by induction on $k = 0, 1 \ldots$ where k is the iteration of value iteration.
- We know $\lim_{k\to\infty} V^{k}(b) = V^{*}(b)$.
- Define $W^{k}_{l}(b(1)) = V^{k}_{l}(b(1)) - V^{k}_{l-1}(b(1))$

[38] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[39]

- If $b(1) \in \mathscr{S}_{l-1}$, we use the Bellman eq. to obtain:
- $$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} \geq \sum_o \mathbb{P}^o_{b(1)} \Big(V^*_{l-1}(b^o(1)) - V^*_{l-2}(b^o(1)) \Big)$$
- **We show that LHS is non-decreasing in l and RHS is non-increasing in l.**
- LHS is non-decreasing by definition of reward function.
- We show that RHS is non-increasing by induction on $k = 0, 1 \ldots$ where k is the iteration of value iteration.
- We know $\lim_{k \to \infty} V^k(b) = V^*(b)$.
- Define $W^k_l(b(1)) = V^k_l(b(1)) - V^k_{l-1}(b(1))$

[39] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[40]

Proof.

$W_l^0(b(1)) = 0 \ \forall l$. Assume $W_{l-1}^{k-1}(b(1)) - W_l^{k-1}(b(1)) \geq 0$. $\square$

[40] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[41]

Proof.

$W_l^0(b(1)) = 0 \ \forall l$. Assume $W_{l-1}^{k-1}(b(1)) - W_l^{k-1}(b(1)) \geq 0$.

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = 2V_{l-1}^k - V_{l-2}^k - V_l^k$$

□

[41] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[42]

Proof.

$W_l^0(b(1)) = 0\ \forall l$. Assume $W_{l-1}^{k-1}(b(1)) - W_l^{k-1}(b(1)) \geq 0$.

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = 2V_{l-1}^k - V_{l-2}^k - V_l^k = 2\mathcal{R}_{b(1)}^{a_{l-1}^k} - \mathcal{R}_{b(1)}^{a_l^k} - \mathcal{R}_{b(1)}^{a_{l-2}^k}$$
$$+ \sum_{o\in\mathcal{O}} \mathbb{P}_{b(1)}^o \Big(2V_{l-1-a_{l-1}^k}^{k-1}(b(1)) - V_{l-a_l^k}^{k-1}(b(1)) - V_{l-2-a_{l-2}^k}^{k-1}(b(1))\Big)$$

Want to show that the above is non-negative. This depends on $a_l^k, a_{l-1}^k, a_{l-2}^k$. $\square$

[42] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[43]

Proof.

$W_l^0(b(1)) = 0 \ \forall l$. Assume $W_{l-1}^{k-1}(b(1)) - W_l^{k-1}(b(1)) \geq 0$.

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = 2V_{l-1}^k - V_{l-2}^k - V_l^k = 2\mathcal{R}_{b(1)}^{a_{l-1}^k} - \mathcal{R}_{b(1)}^{a_l^k} - \mathcal{R}_{b(1)}^{a_{l-2}^k}$$
$$+ \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(2V_{l-1-a_{l-1}^k}^{k-1}(b(1)) - V_{l-a_l^k}^{k-1}(b(1)) - V_{l-2-a_{l-2}^k}^{k-1}(b(1))\Big)$$

Want to show that the above is non-negative. This depends on $a_l^k, a_{l-1}^k, a_{l-2}^k$.

There are four cases to consider:

1. $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k \cap \mathscr{S}_{l-2}^k$
2. $b(1) \in \mathscr{S}_l^k \cap \mathscr{C}_{l-1}^k \cap \mathscr{C}_{l-2}^k$
3. $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k \cap \mathscr{C}_{l-2}^k$
4. $b(1) \in \mathscr{C}_l^k \cap \mathscr{C}_{l-1}^k \cap \mathscr{C}_{l-2}^k$

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[44]

Proof.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k \cap \mathscr{S}_{l-2}^k$, then:

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-2}^{k-1}(b^o(1)) - W_{l-1}^{k-1}(b^o(1)) \Big)$$

which is non-negative by the induction hypothesis. $\square$

[44] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[45]

Proof.

If $b(1) \in \mathscr{S}^k_l \cap \mathscr{S}^k_{l-1} \cap \mathscr{S}^k_{l-2}$, then:

$$W^k_{l-1}(b(1)) - W^k_l(b(1)) = \sum_{o \in \mathcal{O}} \mathbb{P}^o_{b(1)} \Big(W^{k-1}_{l-2}(b^o(1)) - W^{k-1}_{l-1}(b^o(1)) \Big)$$

which is non-negative by the induction hypothesis.

If $b(1) \in \mathscr{S}^k_l \cap \mathscr{C}^k_{l-1} \cap \mathscr{C}^k_{l-2}$, then:

$$W^k_l(b(1)) - W^k_{l-1}(b(1)) = \mathcal{R}^C_{b(1)} - \mathcal{R}^S_{b(1)} + \sum_{o \in \mathcal{O}} \mathbb{P}^o_{b(1)} \Big(W^{k-1}_{l-1}(b^o(1)) \Big)$$

$\square$

[45] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[46]

Proof.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k \cap \mathscr{S}_{l-2}^k$, then:

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-2}^{k-1}(b^o(1)) - W_{l-1}^{k-1}(b^o(1)) \Big)$$

which is non-negative by the induction hypothesis.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{C}_{l-1}^k \cap \mathscr{C}_{l-2}^k$, then:

$$W_l^k(b(1)) - W_{l-1}^k(b(1)) = \mathcal{R}_{b(1)}^C - \mathcal{R}_{b(1)}^S + \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) \Big)$$

Bellman eq. implies, if $b(1) \in \mathscr{C}_{l-1}$, then:

$$\mathcal{R}_{b(1)}^C - \mathcal{R}_{b(1)}^S + \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) \Big) \geq 0$$

$\square$

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[47]

Proof.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k \cap \mathscr{C}_{l-2}^k$, then:

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = \mathcal{R}_{b(1)}^S - \mathcal{R}_{b(1)}^C - \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) \Big)$$

$\square$

[47] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$ [48]

Proof.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k \cap \mathscr{C}_{l-2}^k$, then:

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = \mathcal{R}_{b(1)}^S - \mathcal{R}_{b(1)}^C - \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) \Big)$$

Bellman eq. implies, if $b(1) \in \mathscr{S}_{l-1}^k$, then:

$$\mathcal{R}_{b(1)}^S - \mathcal{R}_{b(1)}^C - \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) \Big) \geq 0$$

$\square$

[48] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[49]

Proof.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k \cap \mathscr{C}_{l-2}^k$, then:

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = \mathcal{R}_{b(1)}^S - \mathcal{R}_{b(1)}^C - \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) \Big)$$

Bellman eq. implies, if $b(1) \in \mathscr{S}_{l-1}^k$, then:

$$\mathcal{R}_{b(1)}^S - \mathcal{R}_{b(1)}^C - \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) \Big) \geq 0$$

If $b(1) \in \mathscr{C}_l^k \cap \mathscr{C}_{l-1}^k \cap \mathscr{C}_{l-2}^k$, then:

$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) - W_l^{k-1}(b^o(1)) \Big)$$

which is non-negative by the induction hypothesis. □

[49] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[50]

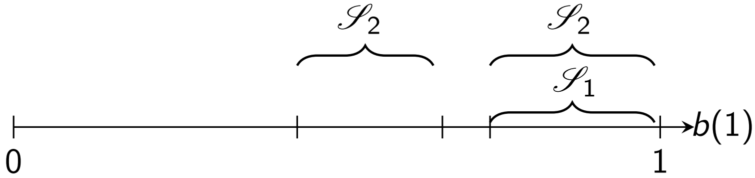
Hence, we have shown that W_l^k is non-increasing in l.

It follows that $b(1) \in \mathscr{S}_{l-1} \implies b(1) \in \mathscr{S}_l$.

[50] T. Nakai. "The problem of optimal stopping in a partially observable Markov chain". In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[51]

$\mathscr{S}_1 \subseteq \mathscr{S}_2$ still allows:



[51] T. Nakai. “The problem of optimal stopping in a partially observable Markov chain”. In: *Journal of Optimization Theory and Applications* 45.3 (1985), pp. 425–442. ISSN: 1573-2878. DOI: 10.1007/BF00938445. URL: https://doi.org/10.1007/BF00938445.

Proofs: Necessary Condition, Total Positivity of Order 2[52]

- A row-stochastic matrix is totally positive of order 2 (TP2) if:
  - The rows of the matrix are stochastically monotone
  - Equivalently, all second-order minors are non-negative.
- Example:

$$A = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{bmatrix} \tag{11}$$

There are $\binom{3}{2}^2$ second-order minors:

$$det \begin{bmatrix} 0.3 & 0.5 \\ 0.2 & 0.4 \end{bmatrix} = 0.02, \quad det \begin{bmatrix} 0.2 & 0.4 \\ 0.1 & 0.2 \end{bmatrix} = 0, \ldots\text{etc.} \tag{12}$$

Since all minors are non-negative, the matrix is TP2

[52] Samuel Karlin. "Total positivity, absorption probabilities and applications". In: *Transactions of the American Mathematical Society* 111 (1964).

Proofs: Necessary Condition, Total Positivity of Order 2[53]

- A row-stochastic matrix is totally positive of order 2 (TP2) if:
  - The rows of the matrix are stochastically monotone
  - Equivalently, all second-order minors are non-negative.

- Example:

$$A = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{bmatrix} \tag{13}$$

There are $\binom{3}{2}^2$ second-order minors:

$$det \begin{bmatrix} 0.3 & 0.5 \\ 0.2 & 0.4 \end{bmatrix} = 0.02, \quad det \begin{bmatrix} 0.2 & 0.4 \\ 0.1 & 0.2 \end{bmatrix} = 0, \ldots \text{etc.} \tag{14}$$

Since all minors are non-negative, the matrix is TP2

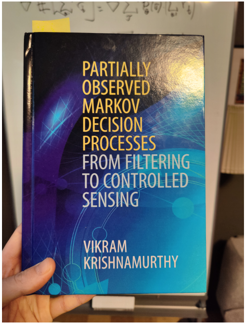
[53] Samuel Karlin. "Total positivity, absorption probabilities and applications". In: *Transactions of the American Mathematical Society* 111 (1964).

Proofs: Monotone belief update[55]

Theorem (Monotone belief update)

Given two beliefs $b_1(1) \geq b_2(1)$, if the transition probabilities and the observation probabilities are TP2, then $b^o_{a,1}(1) \geq b^o_{a,2}(1)$, where $b^o_{a,1}(1)$ and $b^o_{a,2}(1)$ denote the beliefs updated with the Bayesian filter after taking action $a \in \mathcal{A}$ and observing $o \in \mathcal{O}$.

See Theorem 10.3.1 and proof on pp 225,238 in[54]



[54] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316471104.

Proofs: Connected stopping sets $\mathscr{S}_l$[56]

- $\mathscr{S}_l$ is connected if $b(1) \in \mathscr{S}_l, b'(1) \geq b(1) \implies b'(1) \in \mathscr{S}_l$
- If $b(1) \in \mathscr{S}_l$ we use the Bellman eq. to obtain:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + \sum_o \mathbb{P}^o_{b(1)}\Big(V^*_{l-1}(b^o(1)) - V^*_l(b^o(1))\Big) \geq 0$$

- The inequality above should also hold for any $b'(1) \geq b(1)$
- Transition probabilities are TP2 by definition
- We assume observation probabilities are TP2
- It follows that the belief updates are monotone

- Hence, it is sufficient to show that:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^*_{l-1}(b(1)) - V^*_l(b(1))$$

  is weakly increasing in $b(1)$.
- We prove this by induction on k.

[56] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[57]

- $\mathscr{S}_l$ is connected if $b(1) \in \mathscr{S}_l, b'(1) \geq b(1) \implies b'(1) \in \mathscr{S}_l$
- If $b(1) \in \mathscr{S}_l$ we use the Bellman eq. to obtain:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + \sum_o \mathbb{P}^o_{b(1)} \Big(V^*_{l-1}(b^o(1)) - V^*_l(b^o(1)) \Big) \geq 0$$

- The inequality above should also hold for any $b'(1) \geq b(1)$
- Transition probabilities are TP2 by definition
- We assume observation probabilities are TP2
- It follows that the belief updates are monotone

- Hence, it is sufficient to show that:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^*_{l-1}(b(1)) - V^*_l(b(1))$$

  is weakly increasing in $b(1)$.
- We prove this by induction on k.

[57] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[58]

- $\mathscr{S}_l$ is connected if $b(1) \in \mathscr{S}_l, b'(1) \geq b(1) \implies b'(1) \in \mathscr{S}_l$
- If $b(1) \in \mathscr{S}_l$ we use the Bellman eq. to obtain:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + \sum_o \mathbb{P}^o_{b(1)} \Big(V^*_{l-1}(b^o(1)) - V^*_l(b^o(1)) \Big) \geq 0$$

- The inequality above should also hold for any $b'(1) \geq b(1)$
- Transition probabilities are TP2 by definition
- We assume observation probabilities are TP2
- It follows that the belief updates are monotone
- Hence, it is sufficient to show that:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^*_{l-1}(b(1)) - V^*_l(b(1))$$

is weakly increasing in $b(1)$.

- We prove this by induction on k.

[58] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[59]

- $\mathscr{S}_l$ is connected if $b(1) \in \mathscr{S}_l, b'(1) \geq b(1) \implies b'(1) \in \mathscr{S}_l$
- If $b(1) \in \mathscr{S}_l$ we use the Bellman eq. to obtain:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + \sum_o \mathbb{P}^o_{b(1)} \Big(V^*_{l-1}(b^o(1)) - V^*_l(b^o(1)) \Big) \geq 0$$

- The inequality above should also hold for any $b'(1) \geq b(1)$
- Transition probabilities are TP2 by definition
- We assume observation probabilities are TP2
- It follows that the belief updates are monotone
- Hence, it is sufficient to show that:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^*_{l-1}(b(1)) - V^*_l(b(1))$$

is weakly increasing in $b(1)$.

- We prove this by induction on k.

[59] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[60]

- $\mathscr{S}_l$ is connected if $b(1) \in \mathscr{S}_l, b'(1) \geq b(1) \implies b'(1) \in \mathscr{S}_l$
- If $b(1) \in \mathscr{S}_l$ we use the Bellman eq. to obtain:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + \sum_o \mathbb{P}^o_{b(1)}\Big(V^*_{l-1}(b^o(1)) - V^*_l(b^o(1))\Big) \geq 0$$

- The inequality above should also hold for any $b'(1) \geq b(1)$
- Transition probabilities are TP2 by definition
- We assume observation probabilities are TP2
- It follows that the belief updates are monotone

- Hence, it is sufficient to show that:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^*_{l-1}(b(1)) - V^*_l(b(1))$$

  is weakly increasing in $b(1)$.
- We prove this by induction on k.

[60] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[61]

Assume $\mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} + V^{k-1}_{l-1}(b(1)) - V^{k-1}_{l}(b(1))$ is weakly increasing in $b(1)$.

$$\mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} + V^{k}_{l-1}(b(1)) - V^{k}_{l}(b(1)) = \mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} +$$
$$\mathcal{R}^{a^k_{l-1}}_{b(1)} - \mathcal{R}^{a^k_{l}}_{b(1)} + \sum_{o\in\mathcal{O}} \mathbb{P}^{o}_{b(1)}\Big(V^{k-1}_{l-1-a^k_{l-1}}(b^o(1)) - V^{k-1}_{l-a^k_l}(b^o(1))\Big)$$

[61] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[62]

Assume $\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^{k-1}_{l-1}(b(1)) - V^{k-1}_{l}(b(1))$ is weakly increasing in $b(1)$.

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^{k}_{l-1}(b(1)) - V^{k}_{l}(b(1)) = \mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + \mathcal{R}^{a^k_{l-1}}_{b(1)} - \mathcal{R}^{a^k_{l}}_{b(1)} + \sum_{o \in \mathcal{O}} \mathbb{P}^o_{b(1)} \Big(V^{k-1}_{l-1-a^k_{l-1}}(b^o(1)) - V^{k-1}_{l-a^k_l}(b^o(1)) \Big)$$

Want to show that the above is weakly-increasing in $b(1)$. This depends on a^k_l and a^k_{l-1}.

[62] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[63]

Assume $\mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} + V^{k-1}_{l-1}(b(1)) - V^{k-1}_{l}(b(1))$ is weakly increasing in $b(1)$.

$$\mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} + V^{k}_{l-1}(b(1)) - V^{k}_{l}(b(1)) = \mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} +$$

$$\mathcal{R}^{a^k_{l-1}}_{b(1)} - \mathcal{R}^{a^k_{l}}_{b(1)} + \sum_{o \in \mathcal{O}} \mathbb{P}^{o}_{b(1)} \Big(V^{k-1}_{l-1-a^k_{l-1}}(b^o(1)) - V^{k-1}_{l-a^k_l}(b^o(1)) \Big)$$

Want to show that the above is weakly-increasing in $b(1)$. This depends on a^k_l and a^k_{l-1}.

There are three cases to consider:

1. $b(1) \in \mathscr{S}^k_l \cap \mathscr{S}^k_{l-1}$
2. $b(1) \in \mathscr{S}^k_l \cap \mathscr{C}^k_{l-1}$
3. $b(1) \in \mathscr{C}^k_l \cap \mathscr{C}^k_{l-1}$

[63] Kim Hammar and Rolf Stadler. “Intrusion Prevention through Optimal Stopping”. In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[64]

Proof.

If $b(1) \in \mathscr{S}_l \cap \mathscr{S}_{l-1}$, then:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^k_{l-1}(b(1)) - V^k_l(b(1)) =$$

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} \sum_{o \in \mathcal{O}} \mathbb{P}^o_{b(1)} \Big(V^{k-1}_{l-2}(b^o(1)) - V^{k-1}_{l-1}(b^o(1)) \Big)$$

which is weakly increasing in $b(1)$ by the induction hypothesis. $\square$

[64] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[65]

Proof.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{S}_{l-1}^k$, then:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^k_{l-1}(b(1)) - V^k_l(b(1)) =$$
$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} \sum_{o \in \mathcal{O}} \mathbb{P}^o_{b(1)} \Big(V^{k-1}_{l-2}(b^o(1)) - V^{k-1}_{l-1}(b^o(1)) \Big)$$

which is weakly increasing in $b(1)$ by the induction hypothesis.

If $b(1) \in \mathscr{S}_l^k \cap \mathscr{C}_{l-1}^k$, then:

$$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} + V^k_{l-1}(b(1)) - V^k_l(b(1)) =$$
$$\sum_{o \in \mathcal{O}} \mathbb{P}^o_{b(1)} \Big(V^{k-1}_{l-1}(b^o(1)) - V^{k-1}_{l-1}(b^o(1)) \Big) = 0$$

which is trivially weakly increasing in $b(1)$. $\square$

[65] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Connected stopping sets $\mathscr{S}_l$[66]

Proof.

If $b(1) \in \mathscr{C}_l^k \cap \mathscr{C}_{l-1}^k$, then:

$$\mathcal{R}_{b(1)}^S - \mathcal{R}_{b(1)}^C + V_{l-1}^k(b(1)) - V_l^k(b(1)) =$$

$$\mathcal{R}_{b(1)}^S - \mathcal{R}_{b(1)}^C \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(V_{l-1}^{k-1}(b^o(1)) - V_l^{k-1}(b^o(1)) \Big)$$

which is weakly increasing in $b(1)$ by the induction hypothesis. □

Hence, if $b(1) \in \mathscr{S}_l$ and $b'(1) \geq b(1)$ then $b'(1) \in \mathscr{S}_l$.
Therefore, $\mathscr{S}_l$ is connected.

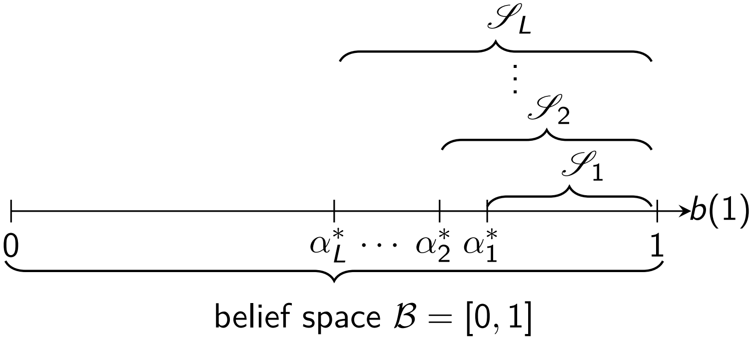
[66] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Proofs: Optimal multi-threshold policy π_l^*[67]

We have shown that:

- $\mathscr{S}_1 = [\alpha_1^*, 1]$
- $\mathscr{S}_l \subseteq \mathscr{S}_{l+1}$
- $\mathscr{S}_l$ is connected (convex) for $l = 1, \ldots, L$

It follows that, $\mathscr{S}_l = [\alpha_l^*, 1]$ and $\alpha_1^* \geq \alpha_2^* \geq \ldots \geq \alpha_L^*$.



[67] Kim Hammar and Rolf Stadler. "Intrusion Prevention through Optimal Stopping". In: (). 2021, https://arxiv.org/abs/2111.00289. arXiv: 2111.00289.

Conclusions & Future Work

- **Conclusions:**
  - We develop a *method* to automatically learn security policies
    - (1) emulation system; (2) system identification; (3) simulation system; (4) reinforcement learning and (5) domain randomization and generalization.
  - We apply the method to an **intrusion prevention use case**
  - We formulate intrusion prevention as a multiple stopping problem
    - We present a POMDP model of the use case
    - We apply the stopping theory to establish structural results of the optimal policy
    - We show numerical results in realistic emulation environment (not included in this presentation)
- **Our research plans:**
  - Extending the model
    - Active attacker: Partially Observed Stochastic Game, Equilibrium analysis
    - Less restrictions on defender

    Scaling up the emulation system:
    - More realistic traffic emulation
    - Non-static infrastructures