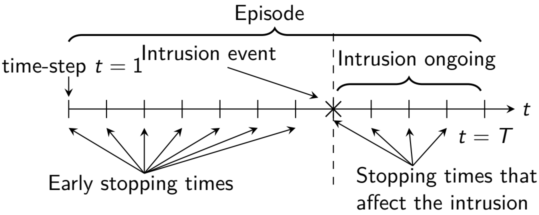
Intrusion Response through Optimal Stopping

New York University - Invited Talk

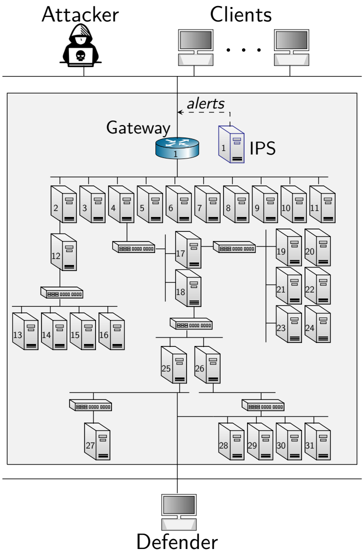
Kim Hammar

kimham@kth.se
Division of Network and Systems Engineering
KTH Royal Institute of Technology

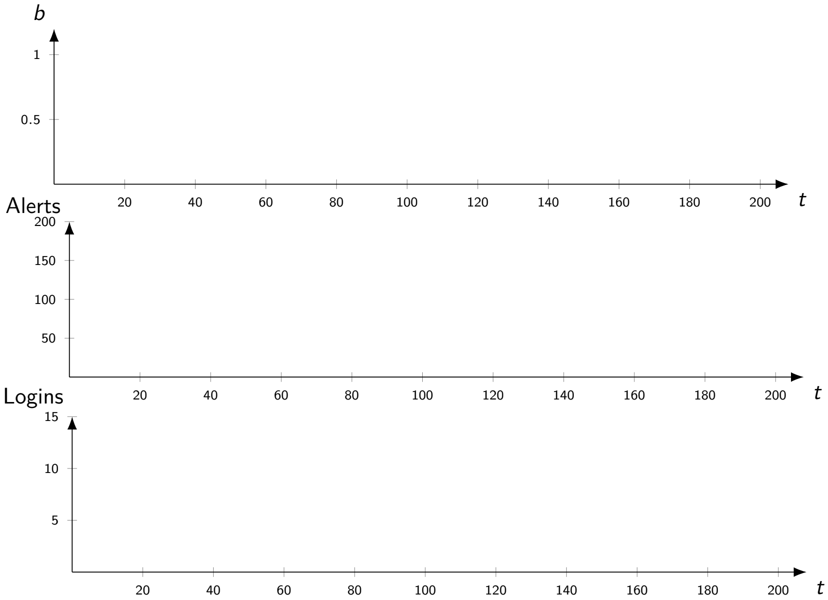


Use Case: Intrusion Response

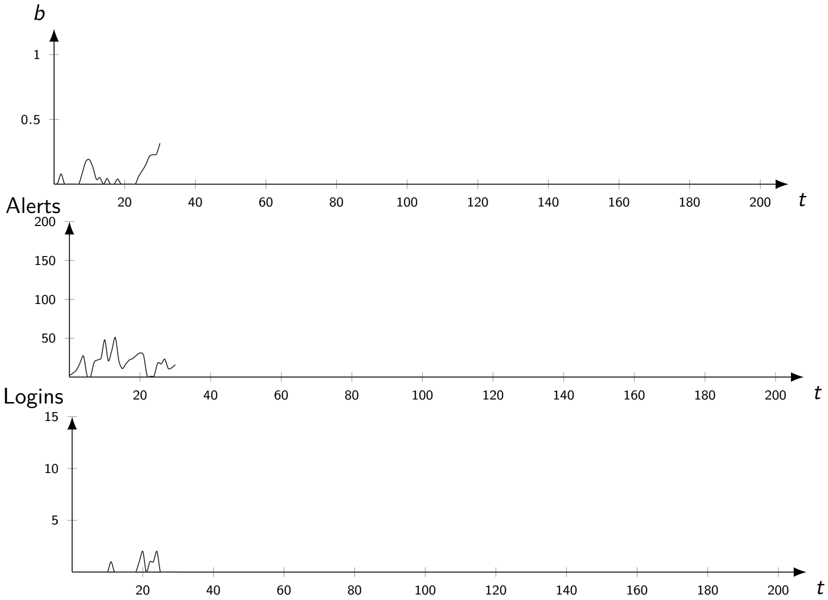
- A **Defender** owns an infrastructure
  - Consists of connected components
  - Components run network services
  - Defender defends the infrastructure by monitoring and active defense
  - Has partial observability
- An **Attacker** seeks to intrude on the infrastructure
  - Has a partial view of the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



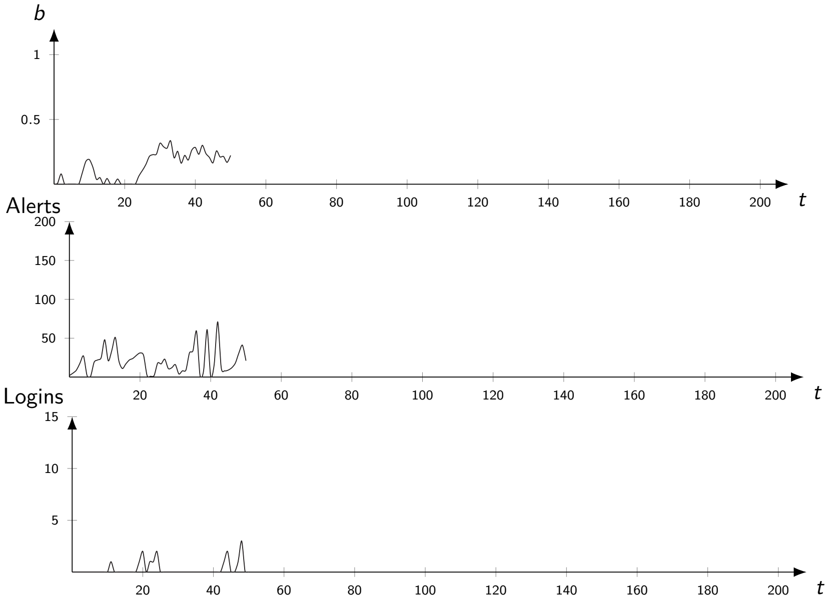
Intrusion Response from the Defender's Perspective



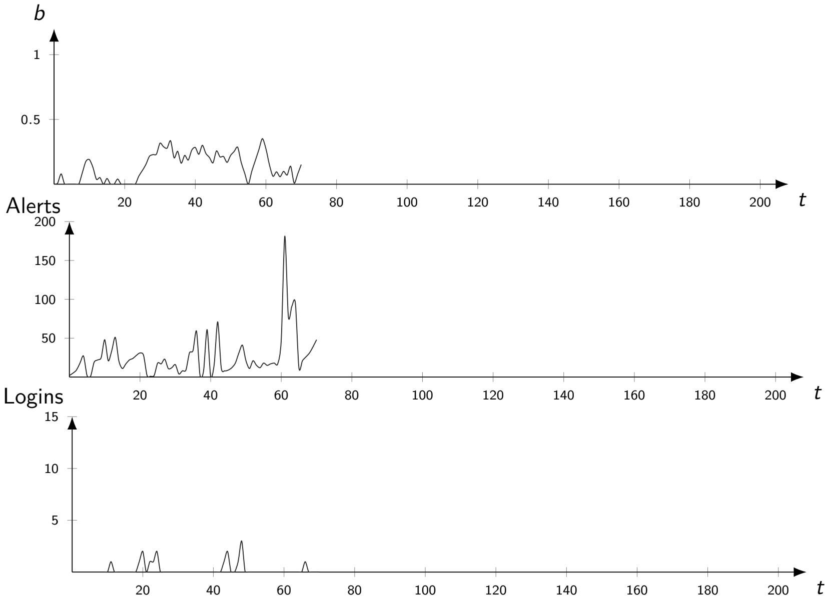
Intrusion Response from the Defender's Perspective



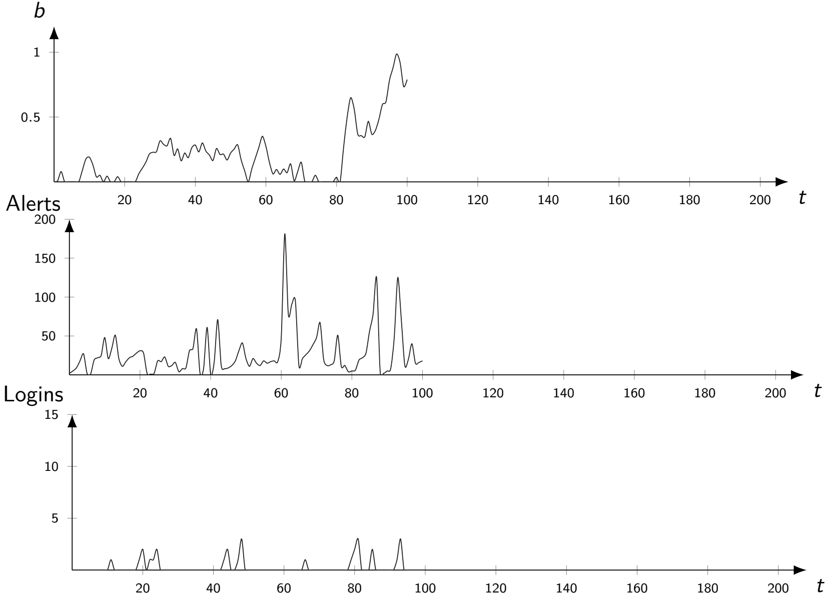
Intrusion Response from the Defender's Perspective



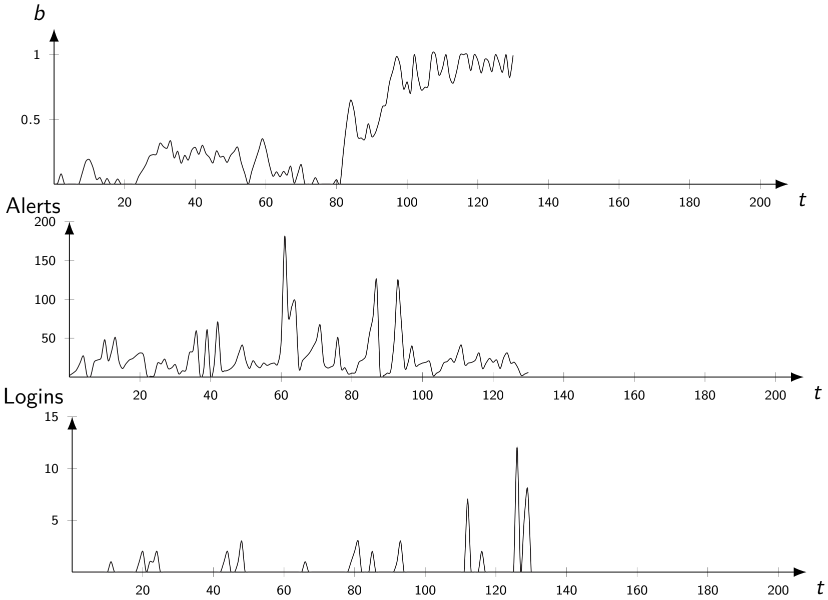
Intrusion Response from the Defender's Perspective



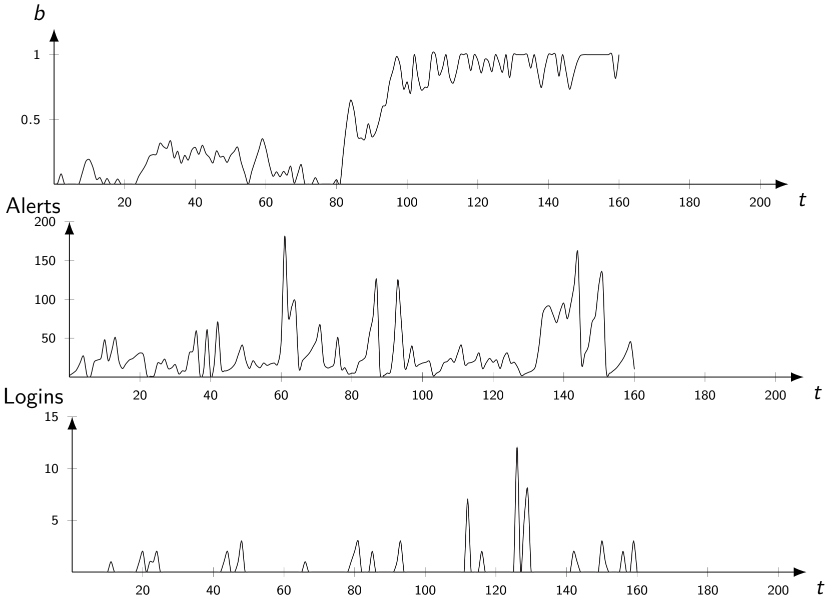
Intrusion Response from the Defender's Perspective



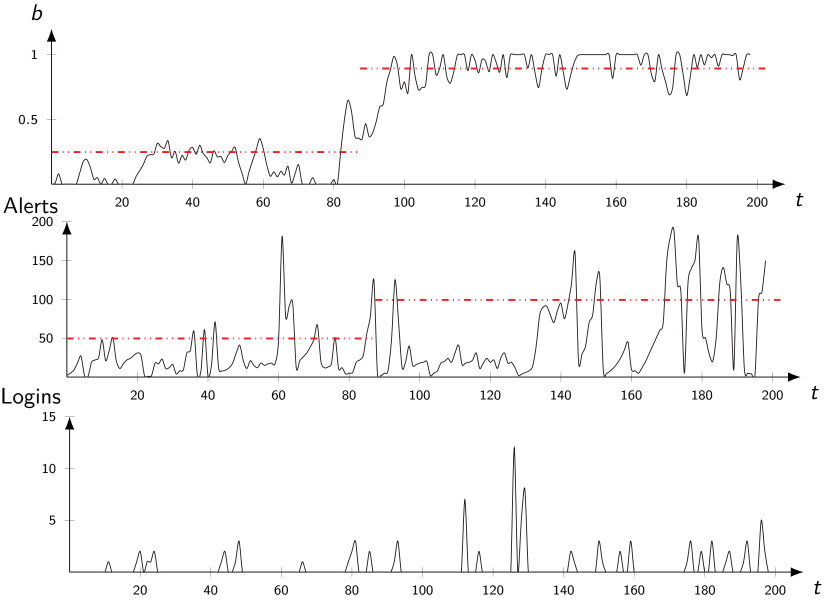
Intrusion Response from the Defender's Perspective



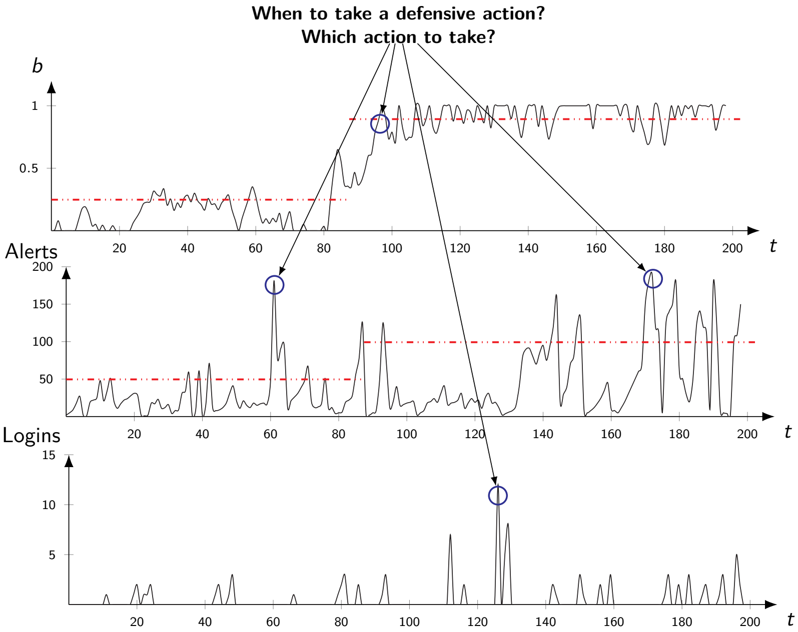
Intrusion Response from the Defender's Perspective



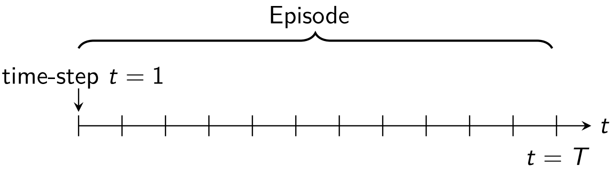
Intrusion Response from the Defender's Perspective

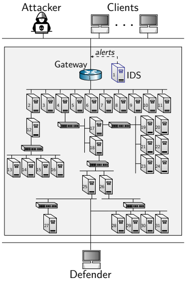


Intrusion Response from the Defender's Perspective



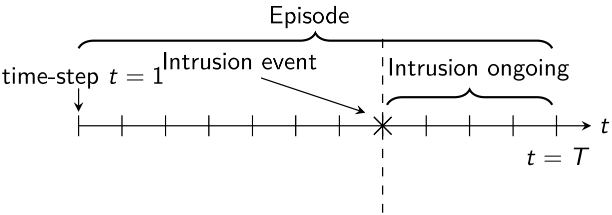
Formulating Intrusion Response as a Stopping Problem

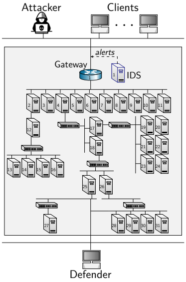




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations,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

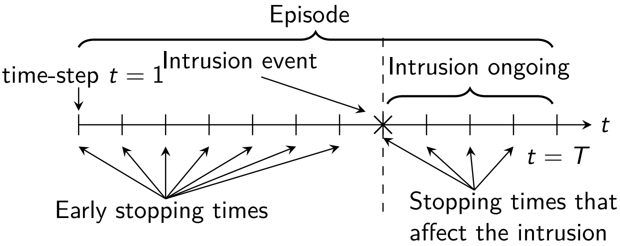
Formulating Intrusion Response as a Stopping Problem

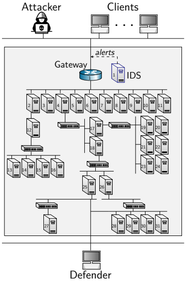




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations,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

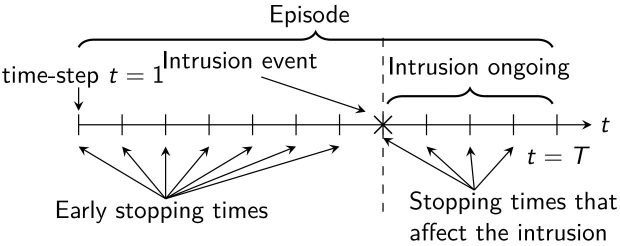
Formulating Intrusion Response as a Stopping Problem

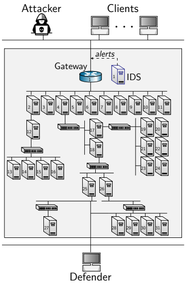




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- Based on the observations,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

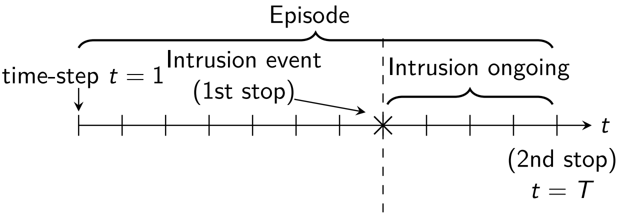
Formulating Intrusion Response as a Stopping Problem

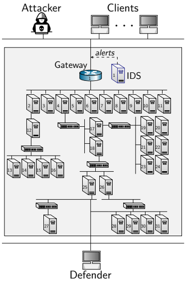




- The system evolves in discrete time-steps.
- Defender observes the infrastructure (IDS, log files, etc.).
- An intrusion occurs at an unknown time.
- The defender can make L stops.
- Each stop is associated with a defensive action
- The final stop shuts down the infrastructure.
- **Based on the observations,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

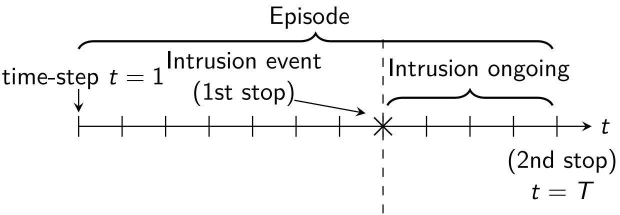
Formulating Network Intrusion as a Stopping Problem

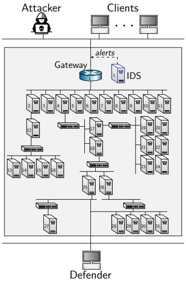




- The system evolves in discrete time-steps.
- The attacker observes the infrastructure (IDS, log files, etc.).
- The first stop action decides when to intrude.
- The attacker can make 2 stops.
- The second stop action terminates the intrusion.
- **Based on the observations & the defender's belief, when is it optimal to stop?**

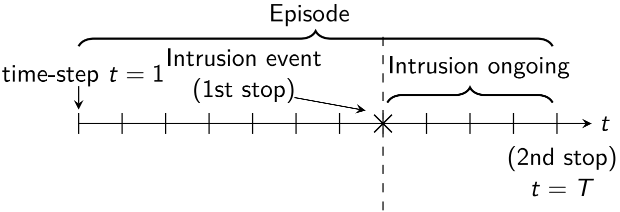
Formulating Network Intrusion as a Stopping Problem

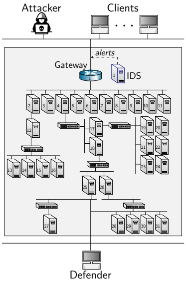




- The system evolves in discrete time-steps.
- The attacker observes the infrastructure (IDS, log files, etc.).
- The first stop action decides when to intrude.
- The attacker can make 2 stops.
- The second stop action terminates the intrusion.
- Based on the observations & the defender's belief, when is it optimal to stop?

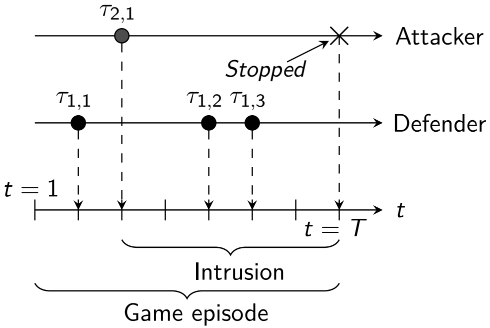
Formulating Network Intrusion as a Stopping Problem





- The system evolves in discrete time-steps.
- The attacker observes the infrastructure (IDS, log files, etc.).
- The first stop action decides when to intrude.
- The attacker can make 2 stops.
- The second stop action terminates the intrusion.
- **Based on the observations & the defender's belief, when is it optimal to stop?**

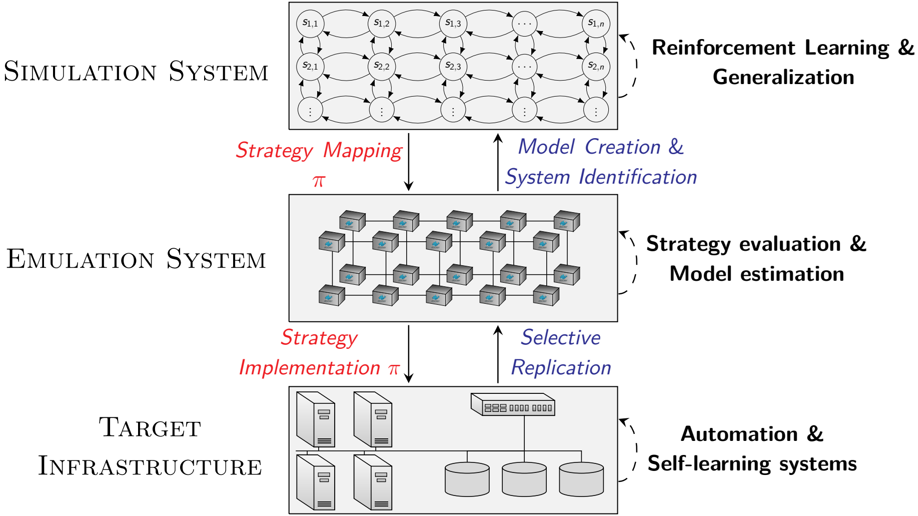
A Dynkin Game Between the Defender and the Attacker[1]



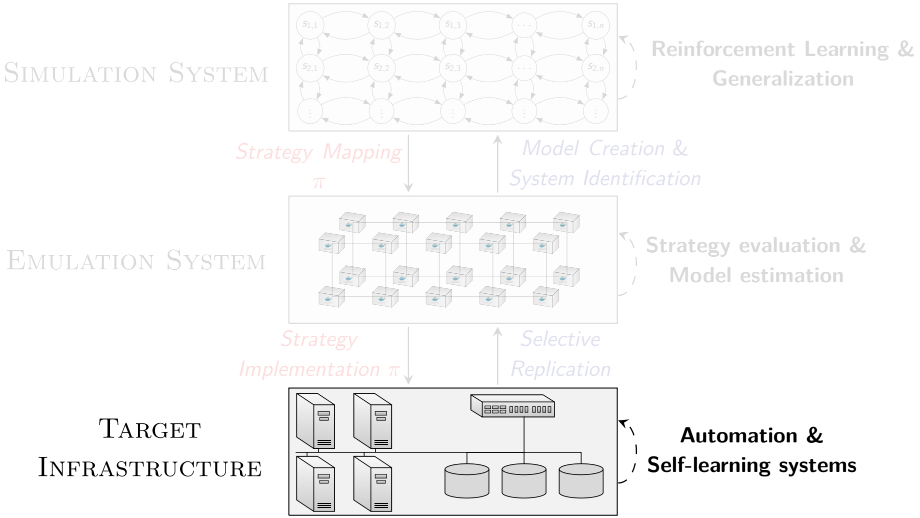
- We formalize the Dynkin game as a zero-sum partially observed one-sided stochastic game.
- The defender is the maximizing player
- The attacker is the minimizing player

[1] E.B Dynkin. "A game-theoretic version of an optimal stopping problem". In: *Dokl. Akad. Nauk SSSR* 385 (1969), pp. 16–19.

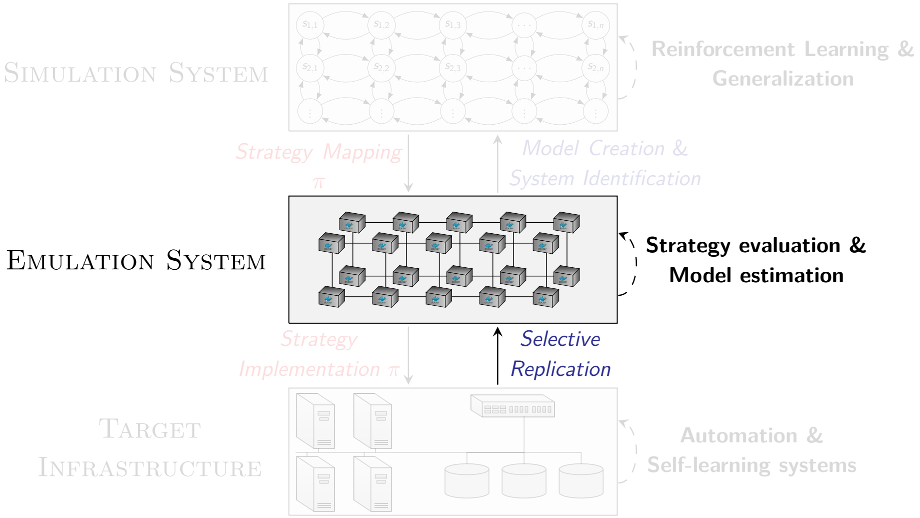
Our Approach for Solving the Dynkin Game



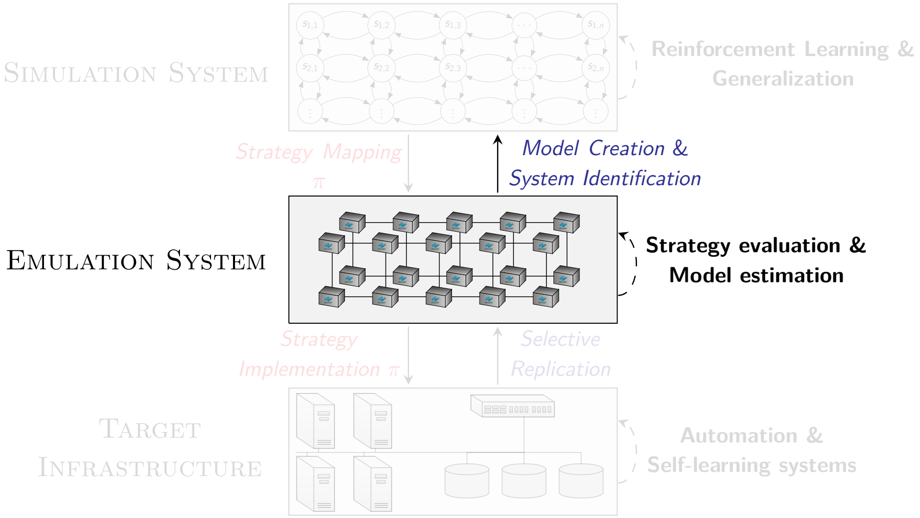
Our Approach for Solving the Dynkin Game



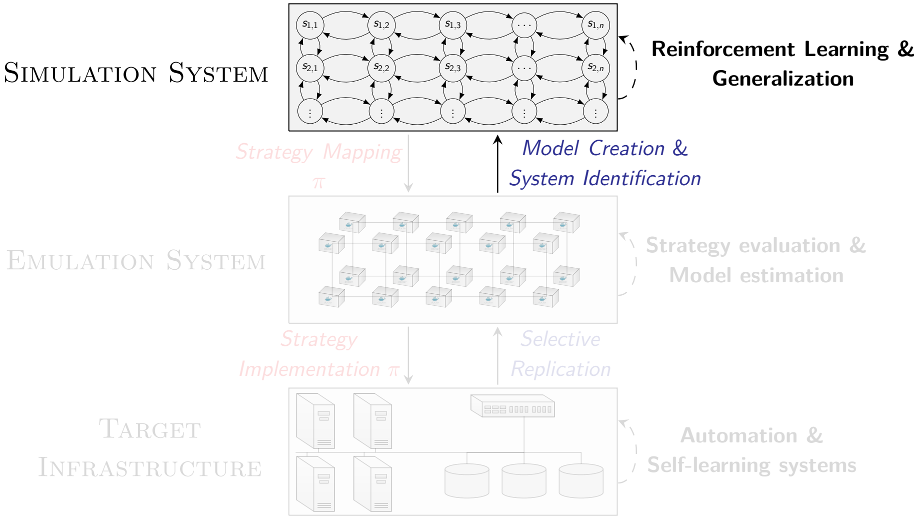
Our Approach for Solving the Dynkin Game



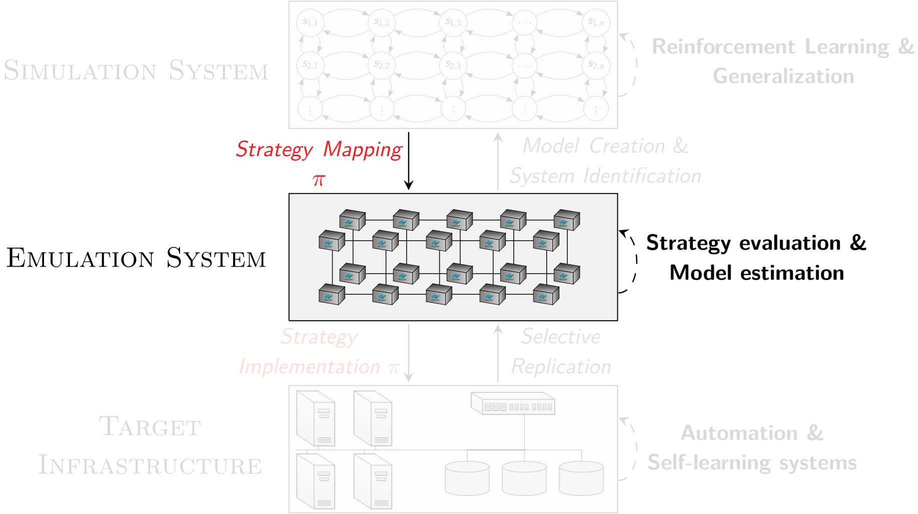
Our Approach for Solving the Dynkin Game



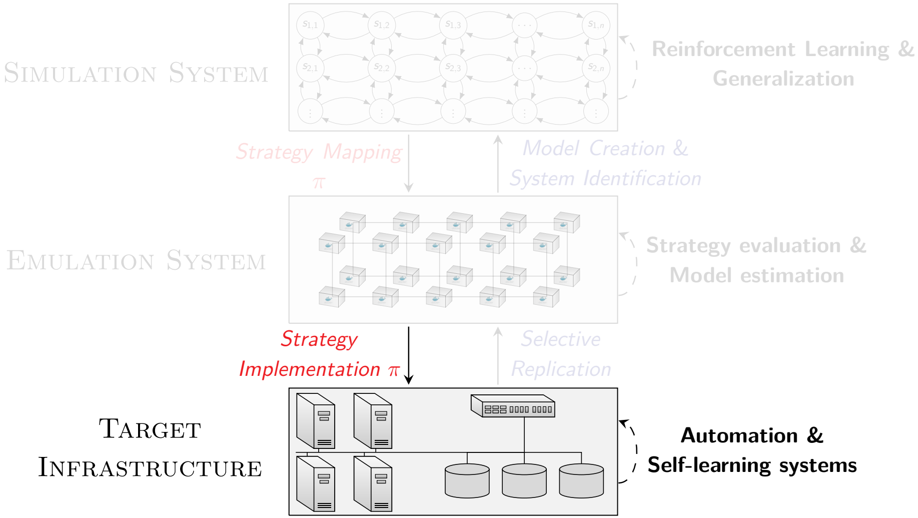
Our Approach for Solving the Dynkin Game



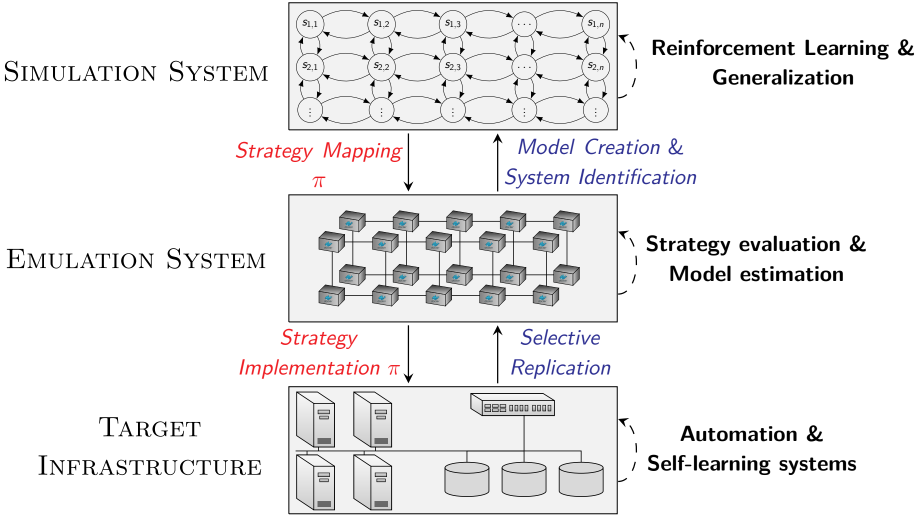
Our Approach for Solving the Dynkin Game



Our Approach for Solving the Dynkin Game



Our Approach for Solving the Dynkin Game



Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - $\mathrm{T\text{-}SPSA}$: A stochastic approximation algorithm to learn π^*
  - $\mathrm{T\text{-}FP}$: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - T-SPSA: A stochastic approximation algorithm to learn π^*
  - T-FP: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - $\mathrm{T\text{-}SPSA}$: A stochastic approximation algorithm to learn π^*
  - $\mathrm{T\text{-}FP}$: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - $\mathrm{T\text{-}SPSA}$: A stochastic approximation algorithm to learn π^*
  - $\mathrm{T\text{-}FP}$: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Outline

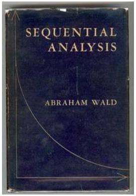
- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - $\mathrm{T\text{-}SPSA}$: A stochastic approximation algorithm to learn π^*
  - $\mathrm{T\text{-}FP}$: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - $\mathrm{T\text{-}SPSA}$: A stochastic approximation algorithm to learn π^*
  - $\mathrm{T\text{-}FP}$: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Optimal Stopping: A Brief History

- **History**:
  - Studied in the 18th century to analyze a gambler's fortune
  - Formalized by Abraham Wald in 1947[2]
  - Since then it has been generalized and developed by (Chow[3], Shiryaev & Kolmogorov[4], Bather[5], Bertsekas[6], etc.)



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

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

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

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

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

The Optimal Stopping Problem

- **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^*:

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

    where $\mathcal{R}^{S}_{ss'}$ & $\mathcal{R}^{C}_{ss'}$ are the stop/continue rewards
  - The $L-l$th **stopping time** τ_l is:

$$\tau_l = \inf\{t : t > \tau_{l-1}, a_t = S\}, \qquad l \in 1, .., L,\ \tau_{L+1} = 0$$

  - τ_l is a random variable from sample space Ω to $\mathbb{N}$, which is dependent on $h_\tau = \rho_1, a_1, o_1, \ldots, a_{\tau-1}, o_\tau$ and independent of $a_\tau, o_{\tau+1}, \ldots$
  - We consider the class of stopping times $\mathcal{T}_t = \{\tau \leq t\} \in \mathcal{F}_k$ where $\mathcal{F}_k$ is the natural filtration on h_t.
- **Solution approaches:** the *Markovian approach* and the *martingale approach*.

The Optimal Stopping Problem

- **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^*:

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

    where $\mathcal{R}^{S}_{ss'}$ & $\mathcal{R}^{C}_{ss'}$ are the stop/continue rewards
  - The $L-l$th **stopping time** τ_l is:

$$\tau_l = \inf\{t : t > \tau_{l-1}, a_t = S\}, \qquad l \in 1, .., L, \ \tau_{L+1} = 0$$

  - τ_l is a random variable from sample space Ω to $\mathbb{N}$, which is dependent on $h_\tau = \rho_1, a_1, o_1, \ldots, a_{\tau-1}, o_\tau$ and independent of $a_\tau, o_{\tau+1}, \ldots$
  - We consider the class of stopping times $\mathcal{T}_t = \{\tau \leq t\} \in \mathcal{F}_k$ where $\mathcal{F}_k$ is the natural filtration on h_t.
- **Solution approaches:** the *Markovian approach* and the *martingale approach*.

Optimal Stopping: Solution Approaches

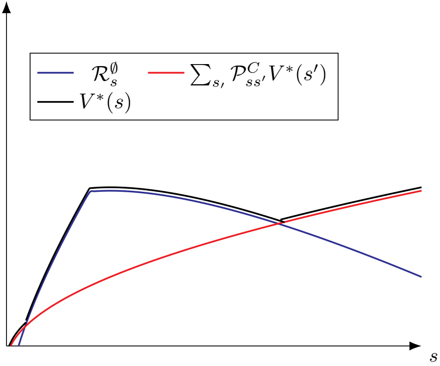
- **The Markovian approach**:
  - Model the problem as a MDP or POMDP
  - A policy π^* that satisfies the Bellman-Wald equation is optimal:

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

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

Optimal Stopping: Solution Approaches

- **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 $R_s^{\emptyset}$.



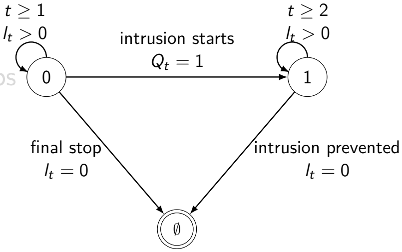
Optimal Stopping: Solution Approaches

- **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[7].

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

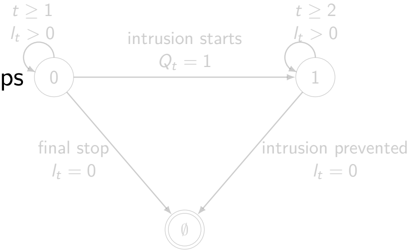
The Defender's Optimal Stopping problem as a POMDP

- **States:**
  - Intrusion state $s_t \in \{0,1\}$, terminal $\emptyset$.
- **Observations:**
  - IDS Alerts weighted by priority o_t, stops remaining $l_t \in \{1,..,L\}$, $f(o_t|s_t)$
- **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:**
  - $\max \mathbb{E}_\pi \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$



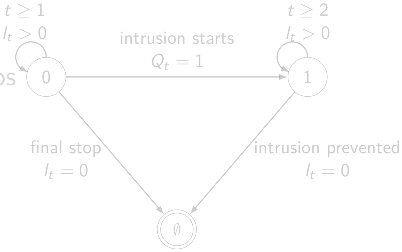
The Defender's Optimal Stopping problem as a POMDP

- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - IDS Alerts weighted by priority o_t, stops remaining $l_t \in \{1, .., L\}$, $f(o_t|s_t)$
- **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:**
  - $\max \mathbb{E}_\pi \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$



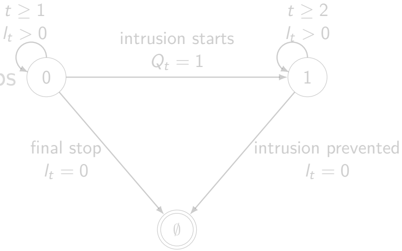
The Defender's Optimal Stopping problem as a POMDP

- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - IDS Alerts weighted by priority o_t, stops remaining $l_t \in \{1, .., L\}$, $f(o_t|s_t)$
- **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:**
  - $\max \mathbb{E}_\pi \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$



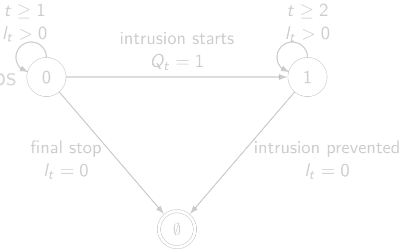
The Defender's Optimal Stopping problem as a POMDP

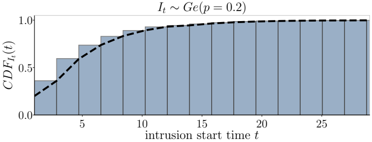
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - IDS Alerts weighted by priority o_t, stops remaining $l_t \in \{1, .., L\}$, $f(o_t|s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false 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:**
  - $\max \mathbb{E}_\pi \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$



The Defender's Optimal Stopping problem as a POMDP

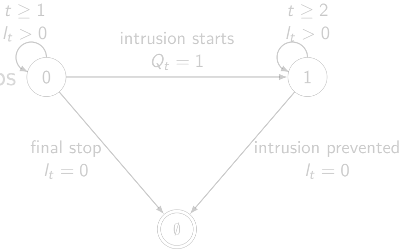
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - IDS Alerts weighted by priority o_t, stops remaining $l_t \in \{1, .., L\}$, $f(o_t|s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false 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:**
  - $\max \mathbb{E}_\pi \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$





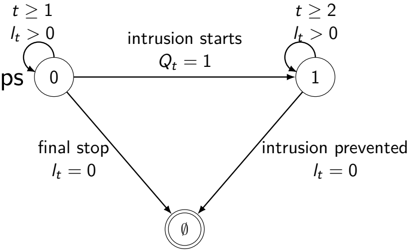
The Defender's Optimal Stopping problem as a POMDP

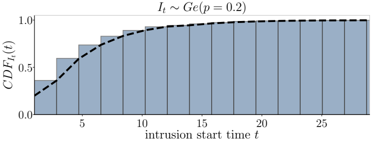
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - IDS Alerts weighted by priority o_t, stops remaining $l_t \in \{1, .., L\}$, $f(o_t|s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false 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:**
  - $\max \mathbb{E}_\pi \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t) \right]$, $T_\emptyset$



The Defender's Optimal Stopping problem as a POMDP

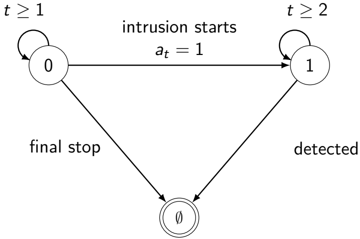
- **States:**
  - Intrusion state $s_t \in \{0, 1\}$, terminal $\emptyset$.
- **Observations:**
  - IDS Alerts weighted by priority o_t, stops remaining $l_t \in \{1, .., L\}$, $f(o_t|s_t)$
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: security and service. Penalty: false 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:**
  - $\max \mathbb{E}_\pi \left[\sum_{t=1}^{T_\emptyset} r(s_t, a_t)\right]$, $T_\emptyset$





The Attacker's Optimal Stopping problem as an MDP

- **States:**
  - Intrusion state $s_t \in \{0,1\}$, terminal $\emptyset$, defender belief $b \in [0,1]$.
- **Actions:**
  - "Stop" (S) and "Continue" (C)
- **Rewards:**
  - Reward: denial of service and intrusion. Penalty: detection
- **Transition probabilities:**
  - Intrusion starts and ends when the attacker takes stop actions
- **Objective and Horizon:**
  - $\max \mathbb{E}_{\pi}\left[\sum_{t=1}^{T_{\emptyset}} r(s_t,a_t)\right],\ T_{\emptyset}$

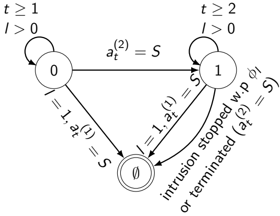


The Dynkin Game as a One-Sided POSG

- **Players:**
  - Player 1 is the defender and player 2 is the attacker. Hence, $\mathcal{N} = \{1, 2\}$.
- **Actions:**
  - $\mathcal{A}_1 = \mathcal{A}_2 = \{S, C\}$.
- **Rewards:**
  - Zero-sum game. Defender maximizes, attacker minimizes.
- **Observability:**
  - The defender has partial observability. The attacker has full observability.
- **Obective functions:**

$$J_1(\pi_1, \pi_2) = \mathbb{E}_{(\pi_1, \pi_2)}\left[\sum_{t=1}^{T} \gamma^{t-1} \mathcal{R}(s_t, \boldsymbol{a}_t)\right] \quad (3)$$

$$J_2(\pi_1, \pi_2) = -J_1(\pi_1, \pi_2) \quad (4)$$



Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - $\mathrm{T\text{-}SPSA}$: A stochastic approximation algorithm to learn π^*
  - $\mathrm{T\text{-}FP}$: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Structural Result: Optimal Multi-Threshold Policy

Theorem

Given the intrusion response POMDP, the following holds:

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

$$\pi_L^*(b(1)) = S \iff b(1) \geq \alpha^* \quad (5)$$

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

$$\pi_l^*(b(1)) = S \iff b(1) \geq \alpha_l^*, \qquad l = 1, \dots, L \quad (6)$$

where α_l^* *is decreasing in* l.

Structural Result: Optimal Multi-Threshold Policy

Theorem

Given the intrusion response 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 defender policy of the form:*

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

3. *If* $L \geq 1$ *and* f_X *is totally positive of order 2 (TP2), there exists L optimal thresholds* $\alpha^*_l \in [0, 1]$ *and an optimal defender 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 response POMDP, the following holds:

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

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

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

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

where α_l^* *is decreasing in* l.

Structural Result: Optimal Multi-Threshold Policy

Theorem

Given the intrusion response POMDP, the following holds:

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

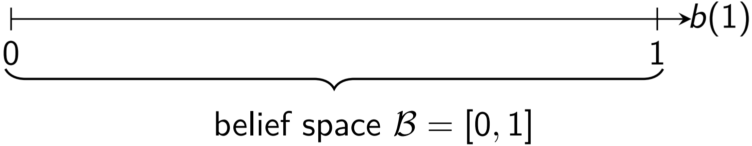
$$\pi^*_L(b(1)) = S \iff b(1) \geq \alpha^* \tag{11}$$

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

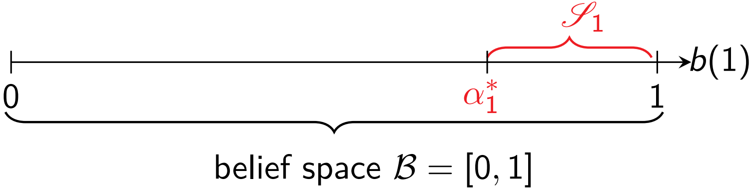
$$\pi^*_l(b(1)) = S \iff b(1) \geq \alpha^*_l, \qquad l = 1, \dots, L \tag{12}$$

where α^*_l *is decreasing in* l.

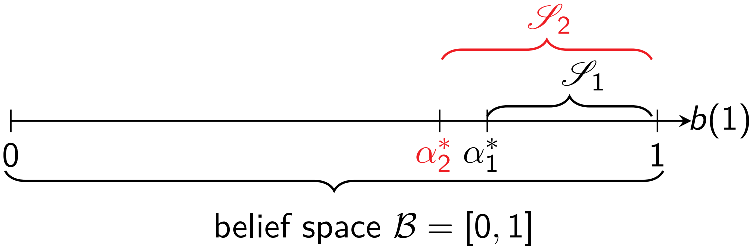
Structural Result: Optimal Multi-Threshold Policy



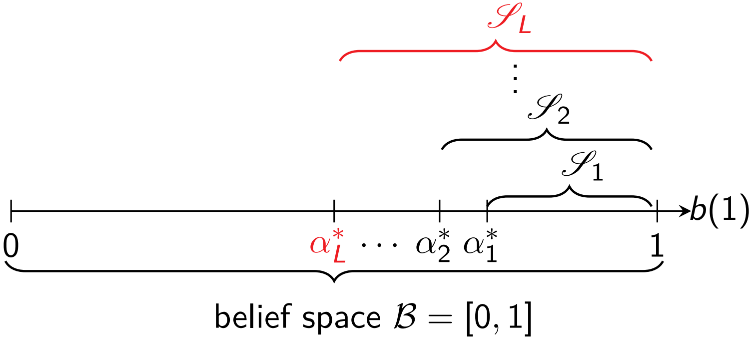
Structural Result: Optimal Multi-Threshold Policy



Structural Result: Optimal Multi-Threshold Policy



Structural Result: Optimal Multi-Threshold Policy

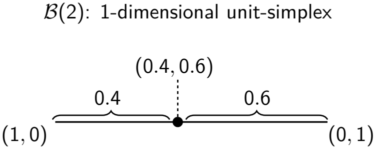


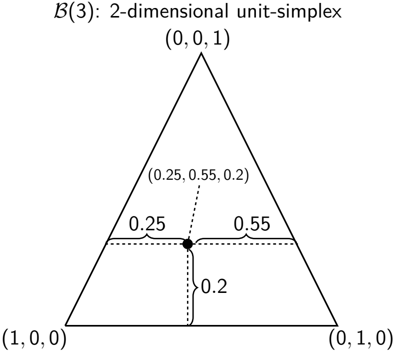
Lemma: $V^*(b)$ is Piece-wise Linear and Convex

Lemma

$V^*(b)$ is piece-wise linear and convex.

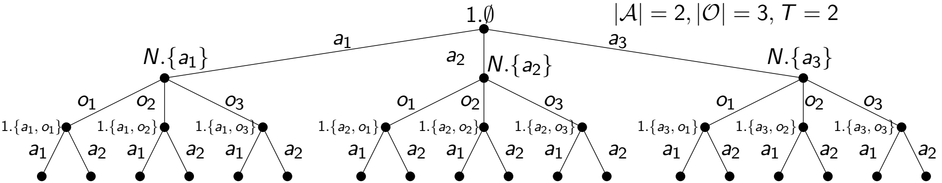
- Belief space $\mathcal{B}$ is the $|\mathcal{S} - 1|$ dimensional unit simplex.
- $|\mathcal{B}| = \infty$, high-dimensional $(|\mathcal{S} - 1|)$ continuous vector
- Infinite set of deterministic policies: $\max_{\pi:\mathcal{B}\rightarrow\mathcal{A}} \mathbb{E}_\pi\left[\sum_t r_t\right]$





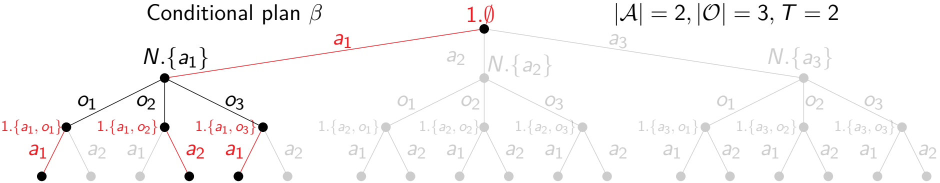
Lemma: $V^*(b)$ is Piece-wise Linear and Convex

- 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**



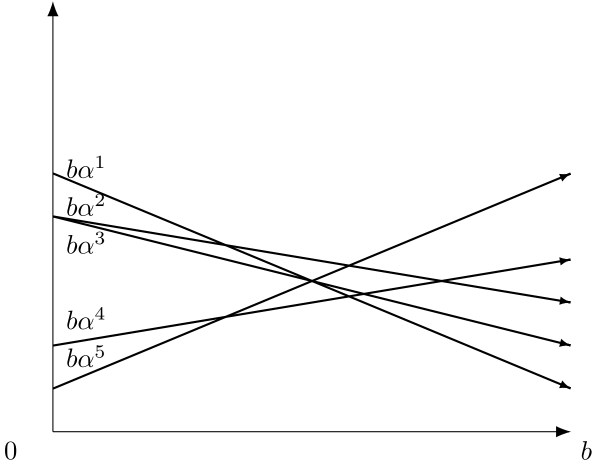
Lemma: $V^*(b)$ is Piece-wise Linear and Convex

- 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**



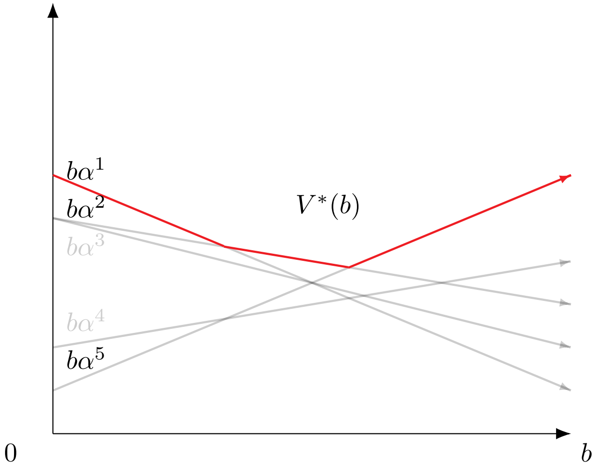
Lemma: $V^*(b)$ is Piece-wise Linear and Convex

- 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[8])



Lemma: $V^*(b)$ is Piece-wise Linear and Convex

- 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[9])



Proofs: $\mathscr{S}_1$ is convex[10]

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

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

[10] 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[11]

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

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

[11] 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[12]

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

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

[12] 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[13]

Proof.

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

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

□

[13] 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[14]

Proof.

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

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

□

[14] 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[15]

Proof.

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

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

□

[15] 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[16]

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. □

[16] 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[17]

Proof.

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

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

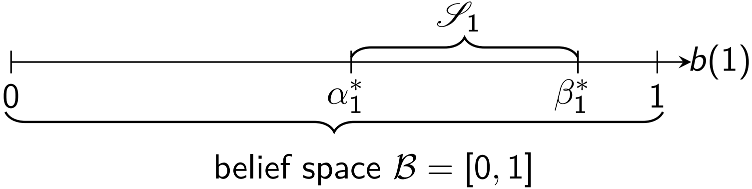
the last inequality is because V^* is optimal. The second-to-last is because there is just a single stop. Hence:

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

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

[17] 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[18]



[18] 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 $L = 1$[19]

- 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:

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

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

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

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

[19] Kim Hammar and Rolf Stadler. "Learning Intrusion Prevention Policies through Optimal Stopping". In: *International Conference on Network and Service Management (CNSM 2021)*. http://dl.ifip.org/db/conf/cnsm/cnsm2021/1570732932.pdf. Izmir, Turkey, 2021.

Proofs: Single-threshold policy is optimal if $L = 1$[20]

- 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:

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

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

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

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

[20] Kim Hammar and Rolf Stadler. "Learning Intrusion Prevention Policies through Optimal Stopping". In: *International Conference on Network and Service Management (CNSM 2021)*. http://dl.ifip.org/db/conf/cnsm/cnsm2021/1570732932.pdf. Izmir, Turkey, 2021.

Proofs: Single-threshold policy is optimal if $L = 1$[21]

- 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:

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

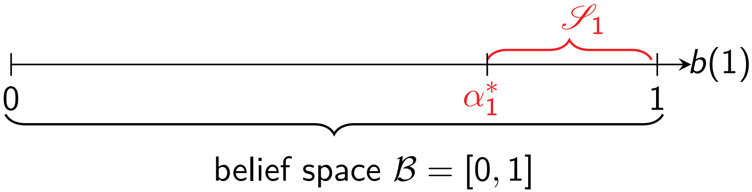
$$= \arg\max_{\{S,C\}} \left[\underbrace{150}_{a=S}, \underbrace{-90 + V^*(1)}_{a=C} \right] = S \quad \text{i.e } \pi^*(1) = \text{Stop}$$

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

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

[21] Kim Hammar and Rolf Stadler. "Learning Intrusion Prevention Policies through Optimal Stopping". In: *International Conference on Network and Service Management (CNSM 2021)*. http://dl.ifip.org/db/conf/cnsm/cnsm2021/1570732932.pdf. Izmir, Turkey, 2021.

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



Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[22]

- We want to show that $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$
- Bellman Equation:

$$\pi^*_{l-1}(b(1)) \in \arg\max_{\{S,C\}} \Bigg[\underbrace{\mathcal{R}^S_{b(1),l-1} + \sum_o \mathbb{P}^o_{b(1)} V^*_{l-2}(b^o(1))}_{\text{Stop}}, \underbrace{\mathcal{R}^C_{b(1),l-1} + \sum_o \mathbb{P}^o_{b(1)} V^*_{l-1}(b^o(1))}_{\text{Continue}} \Bigg]$$

- $\implies$ optimal to stop if:

$$\mathcal{R}^S_{b(1),l-1} - \mathcal{R}^C_{b(1),l-1} \geq \sum_o \mathbb{P}^o_{b(1)} \Big(V^*_{l-1}(b^o(1)) - V^*_{l-2}(b^o(1)) \Big) \quad (13)$$

- Hence, if $b(1) \in \mathscr{S}_{l-1}$, then (13) holds.

[22] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[23]

- We want to show that $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$
- Bellman Equation:

$$\pi^*_{l-1}(b(1)) \in \arg\max_{\{S,C\}} \Bigg[\underbrace{\mathcal{R}^S_{b(1),l-1} + \sum_o \mathbb{P}^o_{b(1)} V^*_{l-2}\big(b^o(1)\big)}_{\text{Stop}}, \underbrace{\mathcal{R}^C_{b(1),l-1} + \sum_o \mathbb{P}^o_{b(1)} V^*_{l-1}\big(b^o(1)\big)}_{\text{Continue}} \Bigg]$$

- $\implies$ optimal to stop if:

$$\mathcal{R}^S_{b(1),l-1} - \mathcal{R}^C_{b(1),l-1} \geq \sum_o \mathbb{P}^o_{b(1)} \Big(V^*_{l-1}\big(b^o(1)\big) - V^*_{l-2}\big(b^o(1)\big) \Big) \quad (13)$$

- Hence, if $b(1) \in \mathscr{S}_{l-1}$, then (13) holds.

[23] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[24]

- We want to show that $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$
- Bellman Equation:

$$\pi^*_{l-1}(b(1)) \in \arg\max_{\{S,C\}} \Bigg[\underbrace{\mathcal{R}^S_{b(1),l-1} + \sum_o \mathbb{P}^o_{b(1)} V^*_{l-2}\big(b^o(1)\big)}_{\text{Stop}}, \underbrace{\mathcal{R}^C_{b(1),l-1} + \sum_o \mathbb{P}^o_{b(1)} V^*_{l-1}\big(b^o(1)\big)}_{\text{Continue}} \Bigg]$$

- $\implies$ optimal to stop if:

$$\mathcal{R}^S_{b(1),l-1} - \mathcal{R}^C_{b(1),l-1} \geq \sum_o \mathbb{P}^o_{b(1)} \Big(V^*_{l-1}\big(b^o(1)\big) - V^*_{l-2}\big(b^o(1)\big) \Big) \tag{13}$$

- Hence, if $b(1) \in \mathscr{S}_{l-1}$, then (13) holds.

[24] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[25]

- $$\mathcal{R}^{S}_{b(1)} - \mathcal{R}^{C}_{b(1)} \geq \sum_{o} \mathbb{P}^{o}_{b(1)}\Big(V^{*}_{l-1}(b^{o}(1)) - V^{*}_{l-2}(b^{o}(1))\Big)$$

- We want to show that $b(1) \in \mathscr{S}_{l-2} \implies b(1) \in \mathscr{S}_{l-1}$.
- Sufficient to show that LHS above 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 $\lim_{k\rightarrow\infty} V^{k}(b) = V^{*}(b)$.
- Define $W^{k}_{l}(b(1)) = V^{k}_{l}(b(1)) - V^{k}_{l-1}(b(1))$

[25] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[26]

- $$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} \geq \sum_o \mathbb{P}^o_{b(1)}\Big(V^*_{l-1}(b^o(1)) - V^*_{l-2}(b^o(1))\Big)$$

- We want to show that $b(1) \in \mathscr{S}_{l-2} \implies b(1) \in \mathscr{S}_{l-1}$.
- Sufficient to show that LHS above 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 $W^k_l(b(1)) = V^k_l(b(1)) - V^k_{l-1}(b(1))$

[26] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[27]

- $$\mathcal{R}^S_{b(1)} - \mathcal{R}^C_{b(1)} \geq \sum_o \mathbb{P}^o_{b(1)}\Big(V^*_{l-1}(b^o(1)) - V^*_{l-2}(b^o(1))\Big)$$

- We want to show that $b(1) \in \mathscr{S}_{l-2} \implies b(1) \in \mathscr{S}_{l-1}$.
- Sufficient to show that LHS above 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 $W^k_l(b(1)) = V^k_l(b(1)) - V^k_{l-1}(b(1))$

[27] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[28]

Proof.

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

[28] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[29]

Proof.

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

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

□

[29] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[30]

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 $a_l^k, a_{l-1}^k, a_{l-2}^k$. □

[30] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[31]

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

The other cases, e.g. $b(1) \in \mathscr{S}_l^k \cap \mathscr{C}_{l-1}^k \cap \mathscr{S}_{l-2}^k$, can be discarded due to the induction assumption. □

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[32]

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. □

[32] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[33]

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:

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

□

[33] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[34]

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:

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

□

Proofs: Nested stopping sets $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[35]

Proof.

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

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

□

[35] 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 $\mathscr{S}_l \subseteq \mathscr{S}_{1+l}$[36]

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:

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

□

[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]

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:

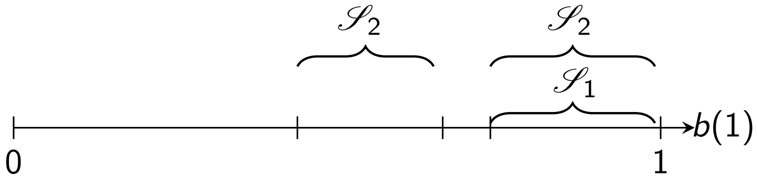
$$W_{l-1}^k(b(1)) - W_l^k(b(1)) = \sum_{o \in \mathcal{O}} \mathbb{P}_{b(1)}^o \Big(W_{l-1}^{k-1}(b^o(1)) - W_l^{k-1}(b^o(1)) \Big)$$

which is non-negative by the induction hypothesis. □

[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.

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

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



We need to show that $\mathscr{S}_l$ is connected, for all $l \in \{1, \ldots, L\}$.

[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: Connected stopping sets $\mathscr{S}_l$[39]

- $\mathscr{S}_l$ 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:

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

- We need to show that the above inequality holds for any $b'(1) \geq b(1)$

[39] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Monotone belief update

Lemma (Monotone belief update)

Given two beliefs $b_1(1) \geq b_2(1)$, if the transition probabilities and the observation probabilities ***are Totally Positive of Order 2 (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[40]

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

Proofs: Necessary Condition, Total Positivity of Order 2[41]

- 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:

$$A = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{bmatrix} \tag{14}$$

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

$$det \begin{bmatrix} 0.3 & 0.5 \\ 0.2 & 0.4 \end{bmatrix} = 0.02, \quad det \begin{bmatrix} 0.2 & 0.4 \\ 0.1 & 0.2 \end{bmatrix} = 0, \text{...etc.} \tag{15}$$

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

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

Proofs: Connected stopping sets $\mathscr{S}_l$[42]

- Since the transition probabilities are TP2 by definition and we assume the observation probabilities are TP2, the condition for showing that the stopping sets are connected reduces to the following.

- Show that the below expression is weakly increasing in $b(1)$.

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

- We prove this by induction on k.

[42] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Connected stopping sets $\mathscr{S}_l$[43]

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 $b(1)$.

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

[43] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Connected stopping sets $\mathscr{S}_l$[44]

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}.

[44] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Connected stopping sets $\mathscr{S}_l$[45]

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

[45] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Connected stopping sets $\mathscr{S}_l$[46]

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. □

[46] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Connected stopping sets $\mathscr{S}_l$[47]

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 $b(1)$. □

[47] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Connected stopping sets $\mathscr{S}_l$[48]

Proof.

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

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

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.

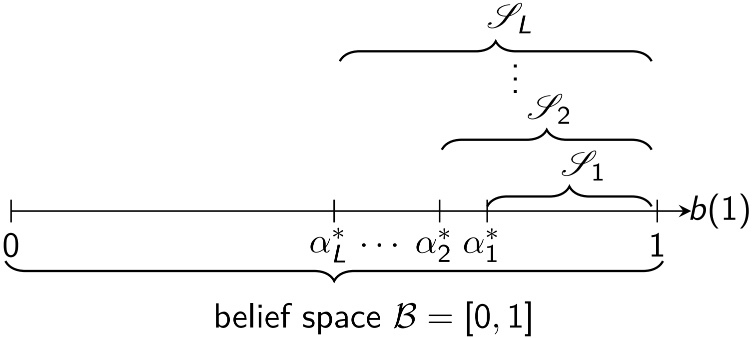
[48] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Proofs: Optimal multi-threshold policy π_l^*[49]

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 $l = 1, \dots, L$

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



[49] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

Structural Result: Best Response Multi-Threshold Attacker Strategy

Theorem

Given the intrusion MDP, the following holds:

1. *Given a defender strategy* $\pi_1 \in \Pi_1$ *where* $\pi_1(S|b(1))$ *is non-decreasing in* $b(1)$ *and* $\pi_1(S|1) = 1$*, then there exist values* $\tilde{\beta}_{0,1}, \tilde{\beta}_{1,1}, \ldots, \tilde{\beta}_{0,L}, \tilde{\beta}_{1,L} \in [0,1]$ *and a best response strategy* $\tilde{\pi}_2 \in B_2(\pi_1)$ *for the attacker that satisfies*

$$\tilde{\pi}_{2,l}(0, b(1)) = C \iff \pi_{1,l}(S|b(1)) \geq \tilde{\beta}_{0,l} \tag{16}$$

$$\tilde{\pi}_{2,l}(1, b(1)) = S \iff \pi_{1,l}(S|b(1)) \geq \tilde{\beta}_{1,l} \tag{17}$$

for $l \in \{1, \ldots, L\}$.

Proof.

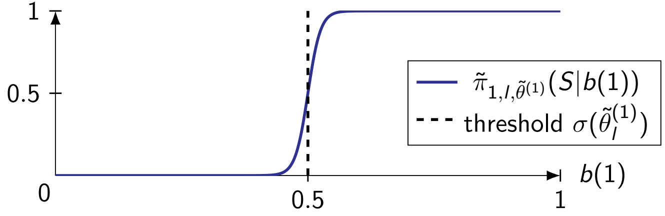
Follows the same idea as the proof for the defender case. See[50]. □

[50] Kim Hammar and Rolf Stadler. *Learning Near-Optimal Intrusion Responses Against Dynamic Attackers*. 2023.

Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - T-SPSA: A stochastic approximation algorithm to learn π^*
  - T-FP: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Threshold-SPSA to Learn Best Responses



A mixed threshold strategy where $\sigma(\tilde{\theta}_l^{(1)})$ is the threshold.

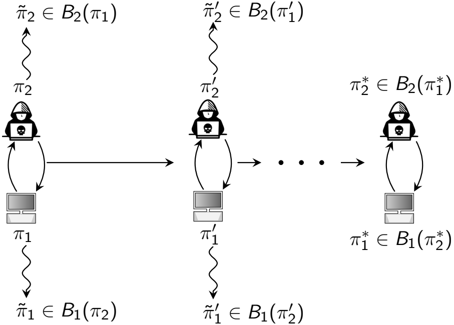
- Parameterizes $\tilde{\pi}_i$ through threshold vectors according to Theorem 1:

$$\varphi(a, b) \triangleq \left(1 + \left(\frac{b(1-\sigma(a))}{\sigma(a)(1-b)}\right)^{-20}\right)^{-1} \tag{18}$$

$$\tilde{\pi}_{i,\tilde{\theta}^{(i)}}(S|b(1)) \triangleq \varphi\left(\tilde{\theta}_l^{(i)}, b(1)\right) \tag{19}$$

- The parameterized strategies are mixed (and differentiable) strategies that approximate threshold strategies.
- Update threshold vectors $\theta^{(i)}$ using SPSA iteratively.

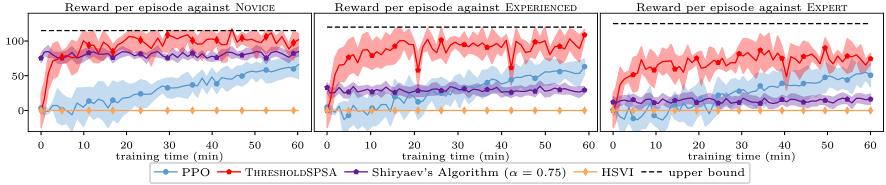
Threshold-Fictitious Play to Approximate an Equilibrium

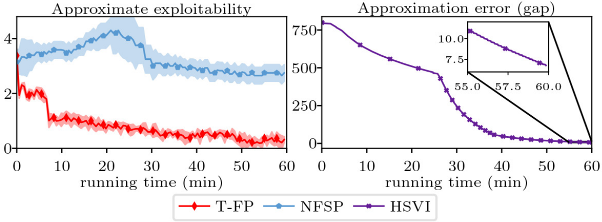


Fictitious play: iterative averaging of best responses.

- Learn best response strategies iteratively through T-SPSA
- Average best responses to approximate the equilibrium

Comparison against State-of-the-art Algorithms





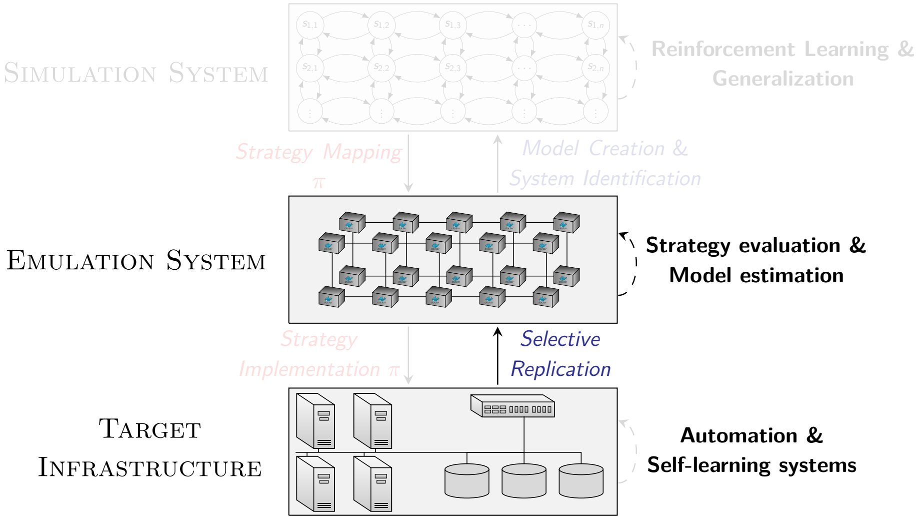
Outline

- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - T-SPSA: A stochastic approximation algorithm to learn π^*
  - T-FP: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Outline

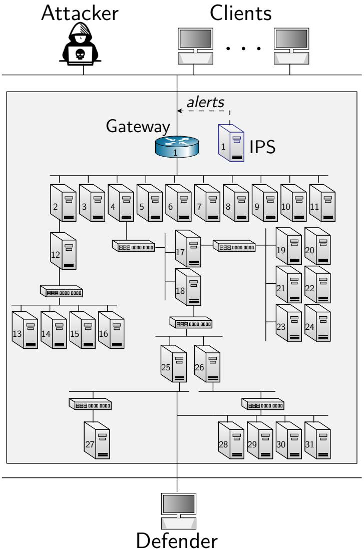
- **Use Case & Approach**
  - Use case: Intrusion response
  - Approach: Optimal stopping
- **Theoretical Background & Formal Model**
  - Optimal stopping problem definition
  - Formulating the Dynkin game as a one-sided POSG
- **Structure of π^***
  - Stopping sets $\mathscr{S}_l$ are connected and nested, $\mathscr{S}_1$ is convex.
  - Existence of multi-threshold best response strategies $\tilde{\pi}_1, \tilde{\pi}_2$.
- **Efficient Algorithms for Learning π^***
  - T-SPSA: A stochastic approximation algorithm to learn π^*
  - T-FP: A Fictitious-play algorithm to approximate (π_1^*, π_2^*)
- **Evaluation Results**
  - Target system, digital twin, system identification, & results
- **Conclusions & Future Work**

Creating a Digital Twin of the Target Infrastructure



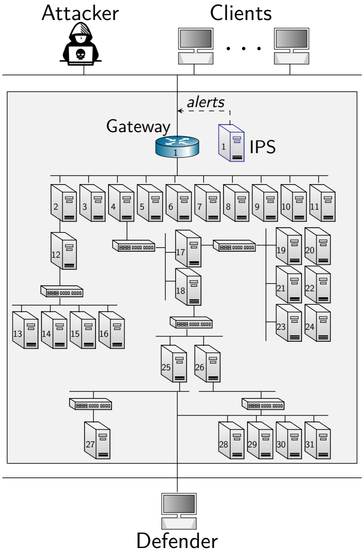
Creating a Digital Twin of the Target Infrastructure

- Emulate **hosts** with docker containers
- Emulate **IPS and vulnerabilities** with software
- Network isolation and **traffic shaping** through NetEm in the Linux kernel
- Enforce **resource constraints** using cgroups.
- Emulate **client arrivals** with Poisson process
- **Internal connections** are full-duplex & loss-less with bit capacities of 1000 Mbit/s
- **External connections** are full-duplex with bit capacities of 100 Mbit/s & 0.1% packet loss in normal operation and random bursts of 1% packet loss



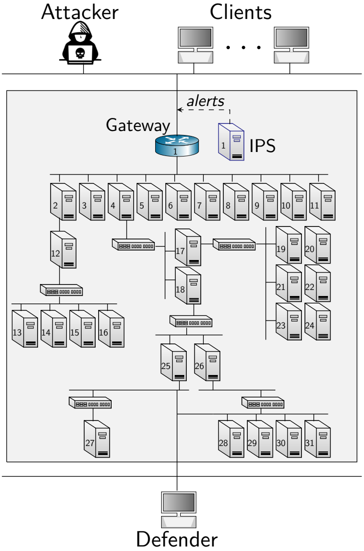
Creating a Digital Twin of the Target Infrastructure

- Emulate **hosts** with docker containers
- Emulate **IPS and vulnerabilities** with software
- Network isolation and **traffic shaping** through NetEm in the Linux kernel
- Enforce **resource constraints** using cgroups.
- Emulate **client arrivals** with Poisson process
- **Internal connections** are full-duplex & loss-less with bit capacities of 1000 Mbit/s
- **External connections** are full-duplex with bit capacities of 100 Mbit/s & 0.1% packet loss in normal operation and random bursts of 1% packet loss



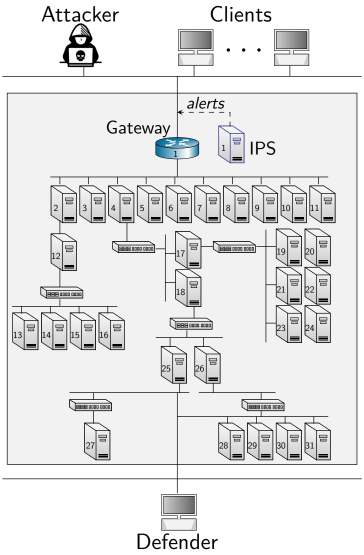
Creating a Digital Twin of the Target Infrastructure

- Emulate **hosts** with docker containers
- Emulate **IPS and vulnerabilities** with software
- Network isolation and **traffic shaping** through NetEm in the Linux kernel
- Enforce **resource constraints** using cgroups.
- Emulate **client arrivals** with Poisson process
- **Internal connections** are full-duplex & loss-less with bit capacities of 1000 Mbit/s
- **External connections** are full-duplex with bit capacities of 100 Mbit/s & 0.1% packet loss in normal operation and random bursts of 1% packet loss



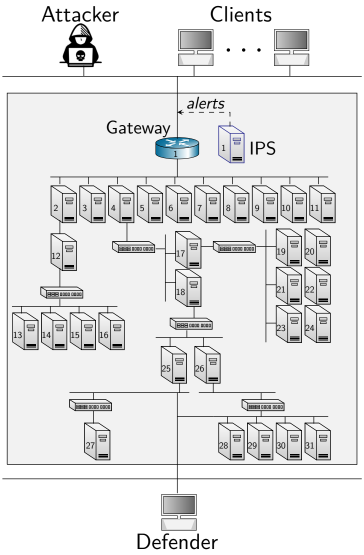
Creating a Digital Twin of the Target Infrastructure

- Emulate **hosts** with docker containers
- Emulate **IPS and vulnerabilities** with software
- Network isolation and **traffic shaping** through NetEm in the Linux kernel
- Enforce **resource constraints** using cgroups.
- Emulate **client arrivals** with Poisson process
- **Internal connections** are full-duplex & loss-less with bit capacities of 1000 Mbit/s
- **External connections** are full-duplex with bit capacities of 100 Mbit/s & 0.1% packet loss in normal operation and random bursts of 1% packet loss



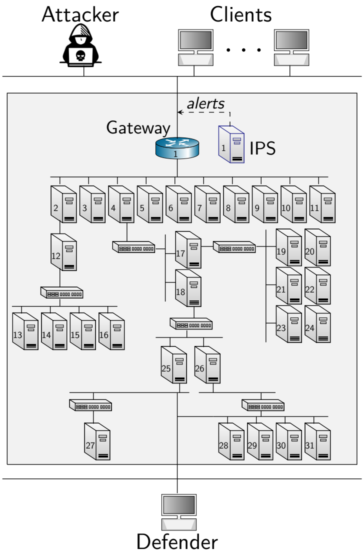
Creating a Digital Twin of the Target Infrastructure

- Emulate **hosts** with docker containers
- Emulate **IPS and vulnerabilities** with software
- Network isolation and **traffic shaping** through NetEm in the Linux kernel
- Enforce **resource constraints** using cgroups.
- Emulate **client arrivals** with Poisson process
- **Internal connections** are full-duplex & loss-less with bit capacities of 1000 Mbit/s
- **External connections** are full-duplex with bit capacities of 100 Mbit/s & 0.1% packet loss in normal operation and random bursts of 1% packet loss

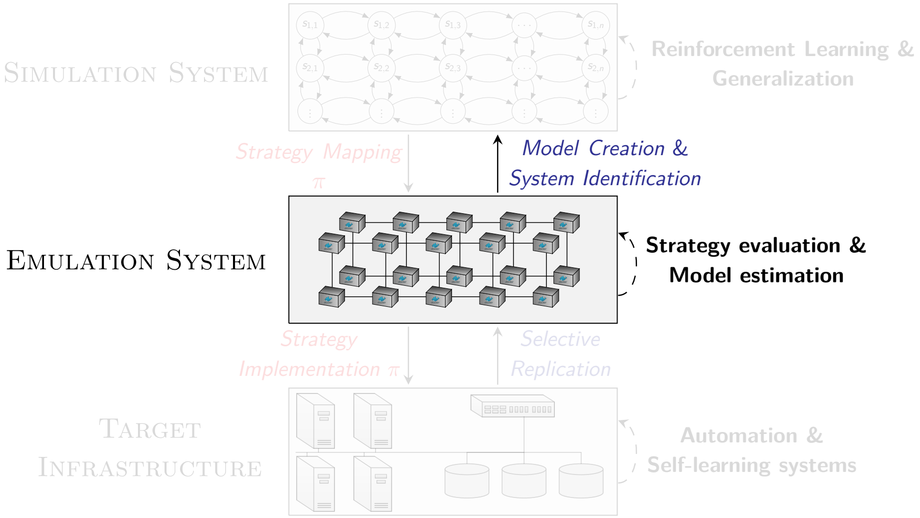


Creating a Digital Twin of the Target Infrastructure

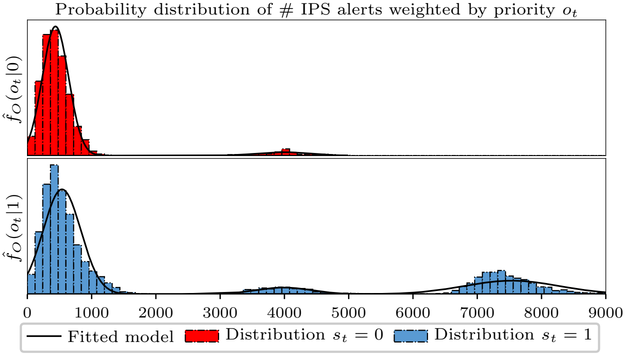
- Emulate **hosts** with docker containers
- Emulate **IPS and vulnerabilities** with software
- Network isolation and **traffic shaping** through NetEm in the Linux kernel
- Enforce **resource constraints** using cgroups.
- Emulate **client arrivals** with Poisson process
- **Internal connections** are full-duplex & loss-less with bit capacities of 1000 Mbit/s
- **External connections** are full-duplex with bit capacities of 100 Mbit/s & 0.1% packet loss in normal operation and random bursts of 1% packet loss



Our Approach for Automated Network Security

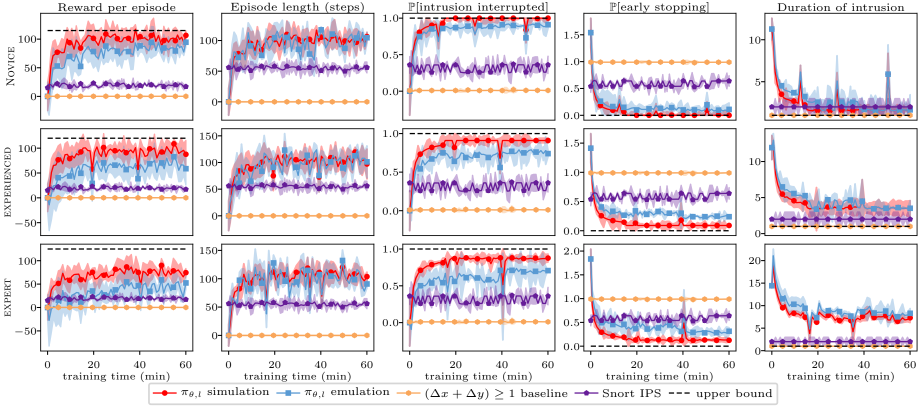


System Identification



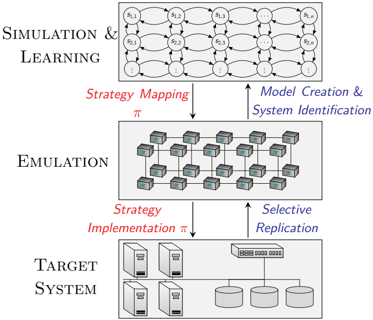
- The distribution f_O of defender observations (system metrics) is unknown.
- We fit a Gaussian mixture distribution $\hat{f}_O$ as an estimate of f_O in the target infrastructure.
- For each state s, we obtain the conditional distribution $\hat{f}_{O|s}$ through expectation-maximization.

Learning Curves in Simulation and Digital Twin

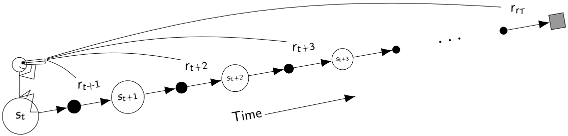


Conclusions

- We develop a *method* to automatically learn security strategies.
- We apply the method to an **intrusion response use case**.
- We design a solution framework guided by the theory of optimal stopping.
- We present several theoretical results on the structure of the optimal solution.
- We show numerical results in a realistic emulation environment.



Current and Future Work



1. **Extend use case**
   - Additional defender actions
   - Utilize SDN controller and NFV-based defenses
   - Increase observation space and attacker model
   - More heterogeneous client population

2. **Extend solution framework**
   - Model-predictive control
   - Rollout-based techniques
   - Extend system identification algorithm

3. **Extend theoretical results**
   - Exploit symmetries and causal structure
   - Utilize theory to improve sample efficiency
   - Decompose solution framework hierarchically