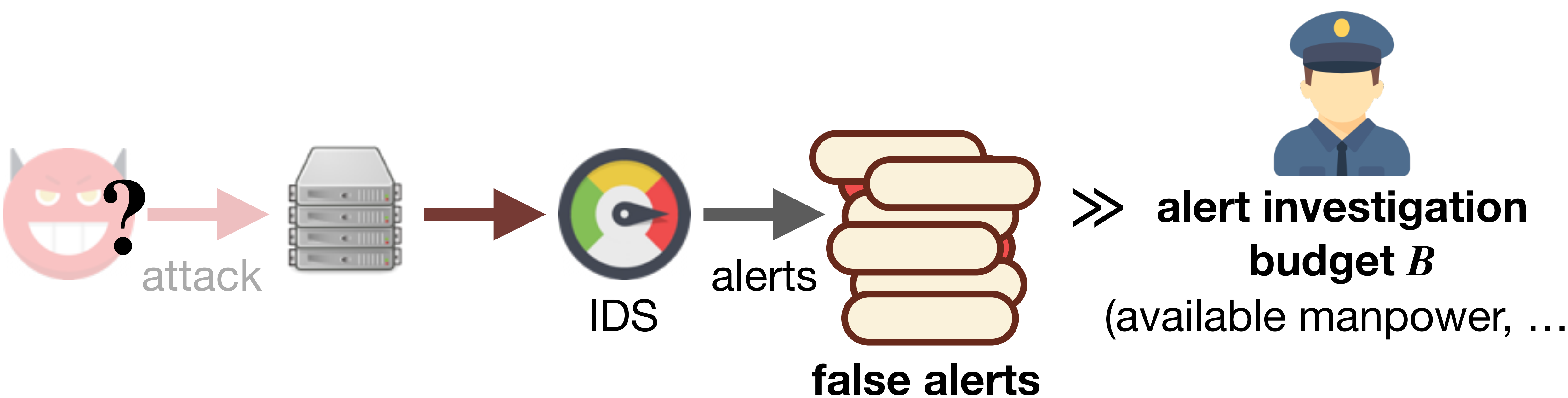
A Game-Theoretic Approach for Alert Prioritization

Aron Laszka, Yevgeniy Vorobeychik, Daniel Fabbri,
Chao Yan, Bradley Malin

VANDERBILT
UNIVERSITY

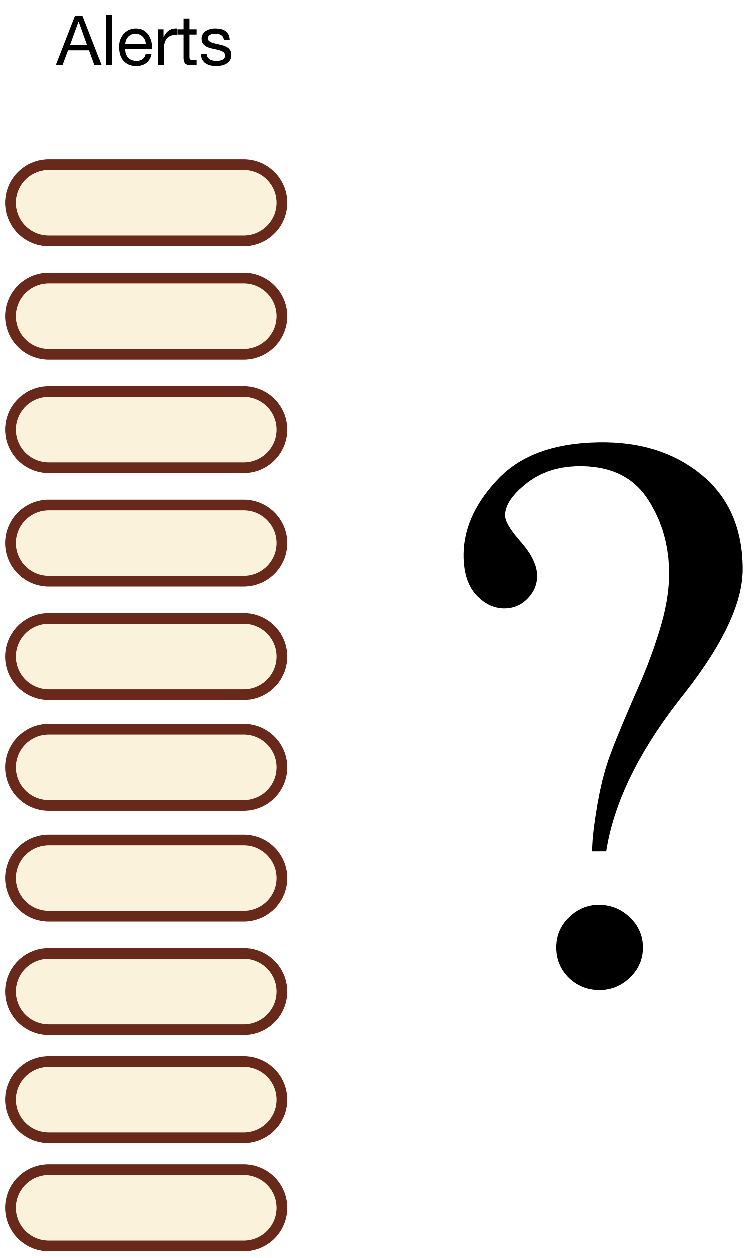
Intrusion Detection

- Detection and mitigation of cyber-attacks is of crucial importance; however, attackers try to stay stealthy
- Intrusion Detection Systems (IDS)
  - generate alerts when they encounter suspicious activity

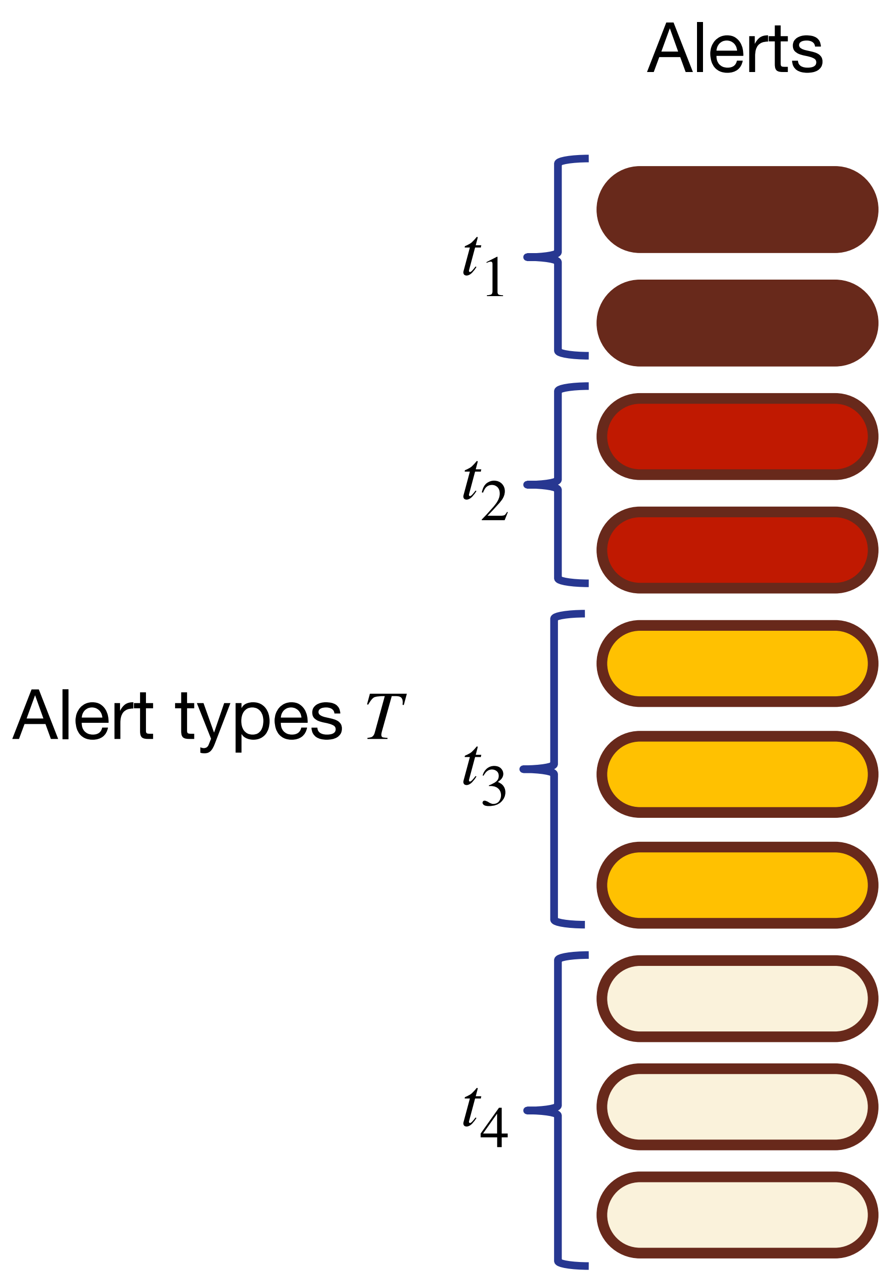


Problem:
Which alerts to investigate?

Alert Prioritization

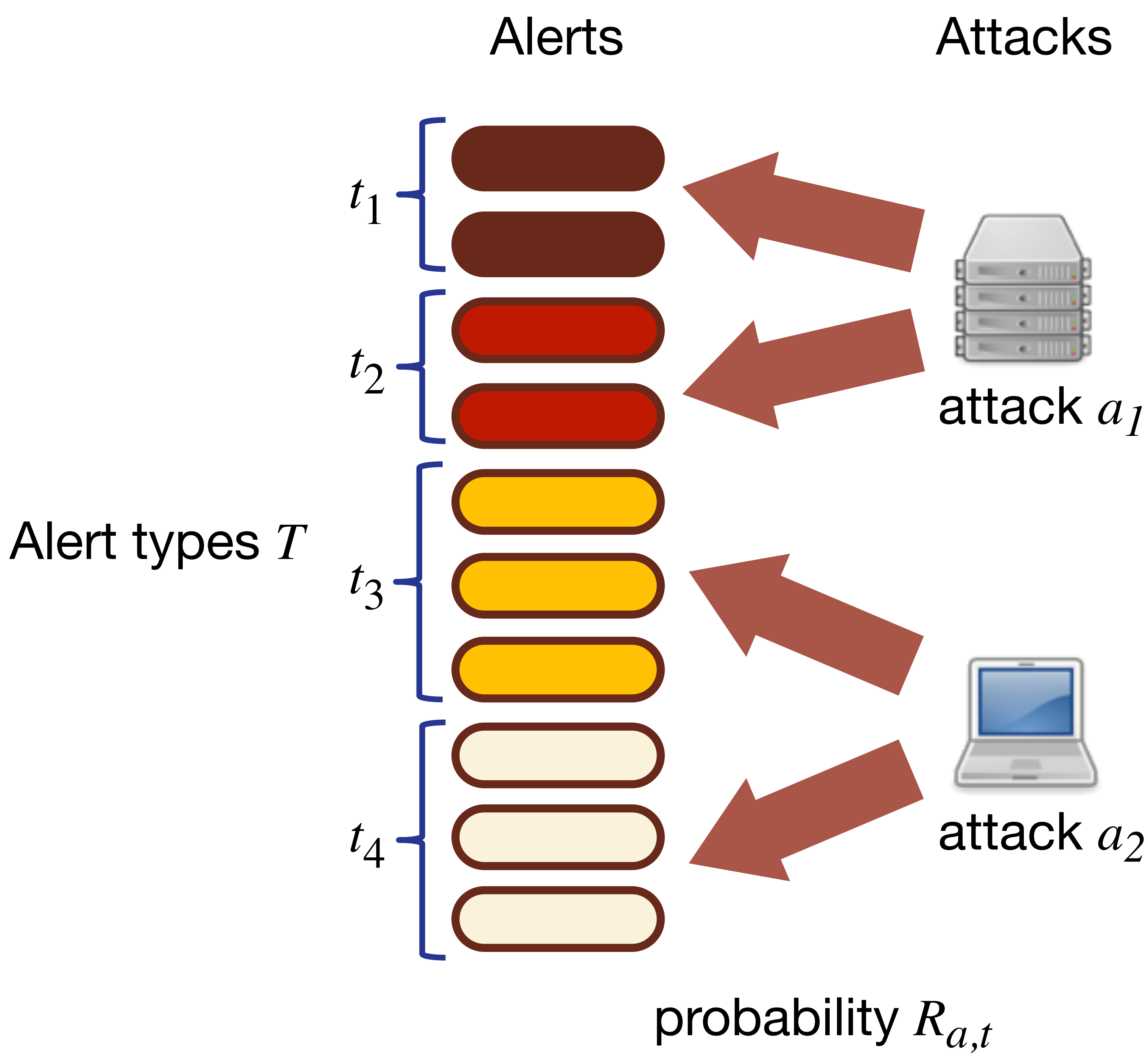


Alert Types



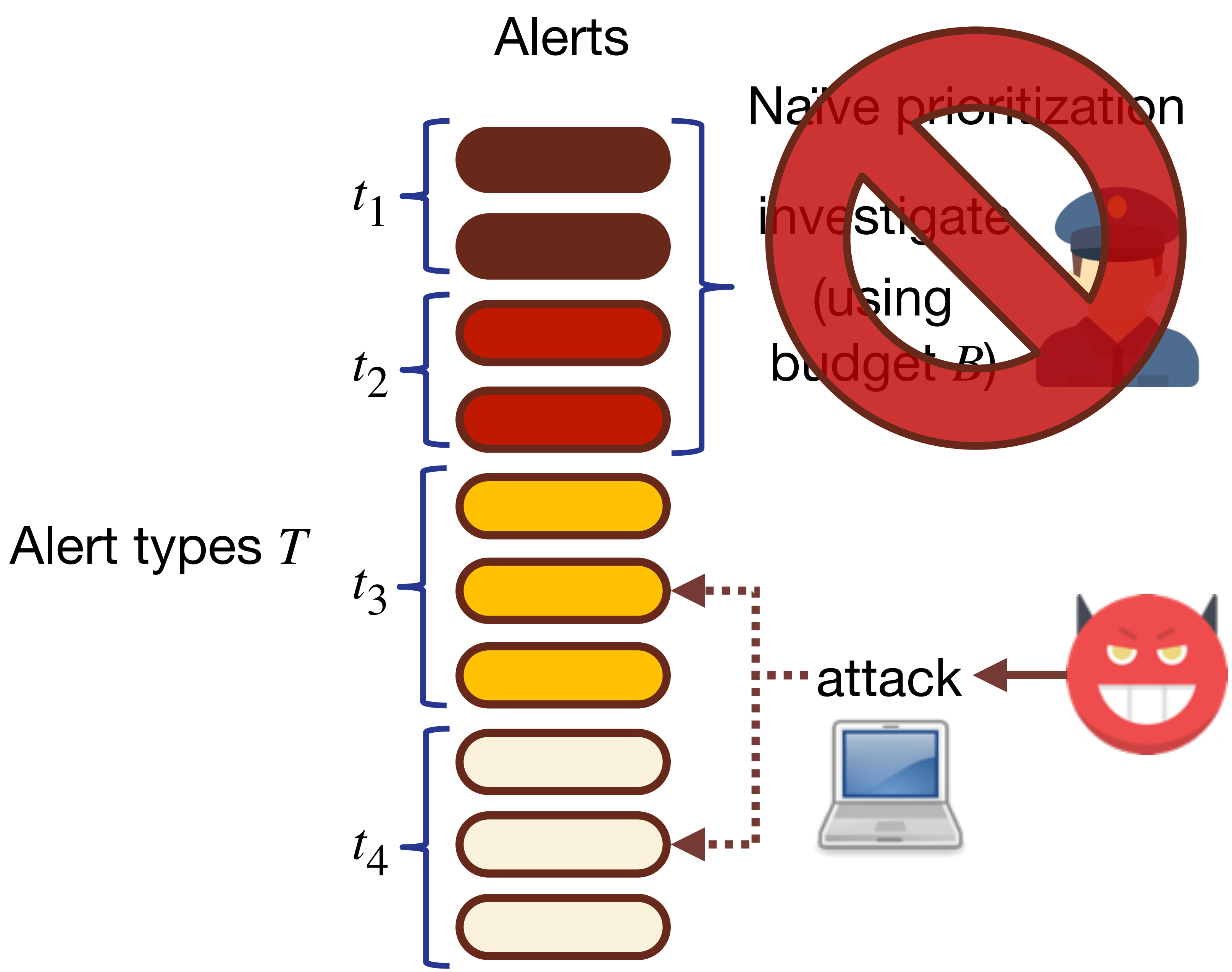
- Alert types T
  - for example, matching different rules in an intrusion detection system (e.g., Snort)
  - before investigating them, alerts of the same type appear equally important
  - cumulative distribution F_t of the number of false alerts of type t is known
- Attacks A
  - for example, targeting certain machines or using certain exploitation techniques
  - impact of attack a is L_a
  - probability of attack a raising an alert of type t is $R_{a,t}$

Alert Types



Alerts
Attacks
t_1
t_2
t_3
t_4
Alert types T
attack a_1
attack a_2
probability $R_{a,t}$

Alert Prioritization Problem



Alert Prioritization Problem

Problem:

What is the optimal probability distribution?

Game-Theoretic Model

- Players

1. Defender: selects an alert prioritization strategy p, which is a probability distribution over possible orderings of T

2. Adversary:
selects an attack a from the set of possible attacks A

- Supposing that the defender uses ordering $\boldsymbol{o} \in T$
  - probability of investigating type k (before exhausting budget B) is

$$PI(\boldsymbol{o}, k) = \sum_{\substack{\boldsymbol{n}: \\ C_{o_k} + \sum_{i=1}^{k} n_i \cdot C_{o_i} \leq B}} \left[\left(F^*_{o_k}(n_k) - F^*_{o_k}(n_k - 1)\right) \cdot \prod_{i=1}^{k-1} \left(F_{o_i}(n_i) - F_{o_i}(n_i - 1)\right) \right]$$

  - probability of investigating attack a (before exhausting budget B) is

$$PD(\boldsymbol{o}, a) = \sum_{\hat{T} \subseteq T} \prod_{t \in \hat{T}} R_{a,t} \prod_{t \in T \setminus \hat{T}} (1 - R_{a,t})\, PI\left(\boldsymbol{o}, \min\{i \,|\, o_i \in \hat{T}\}\right)$$

Optimal Alert Prioritization

- Adversary's gain and defender's loss
  - adversary's expected gain: $EG(\boldsymbol{p}, a) = \sum_{\boldsymbol{o} \in O} p_{\boldsymbol{o}} \cdot (1 - PD(\boldsymbol{o}, a)) \cdot G_a - K_a$
  - defender's expected loss: $EL(\boldsymbol{p}, a) = \sum_{\boldsymbol{o} \in O} p_{\boldsymbol{o}} \cdot (1 - PD(\boldsymbol{o}, a)) \cdot L_a$
- Solution concept: strong Stackelberg equilibrium
  - adversary's best responses: $BR(\boldsymbol{p}) = \operatorname{argmax}_{a \in A} EG(\boldsymbol{p}, a)$
  - **optimal prioritization strategy**: $\min_{\boldsymbol{p}, a \in BR(\boldsymbol{p})} EL(\boldsymbol{p}, a)$

Challenge: finding an optimal probability distribution over a set of exponential size!

Theorem: Finding an optimal alert prioritization strategy is an **NP-hard** problem.

Computing Detection Probabilities

- Probability of detecting an attack

$$PI(\boldsymbol{o}, k) = \sum_{\substack{\boldsymbol{n}: \\ C_{o_k} + \sum_{i=1}^{k} n_i \cdot C_{o_i} \leq B}} \left[\left(F^*_{o_k}(n_k) - F^*_{o_k}(n_k - 1)\right) \cdot \prod_{i=1}^{k-1} \left(F_{o_i}(n_i) - F_{o_i}(n_i - 1)\right) \right]$$

$$PD(\boldsymbol{o}, a) = \sum_{\hat{T} \subseteq T} \prod_{t \in \hat{T}} R_{a,t} \prod_{t \in T \setminus \hat{T}} (1 - R_{a,t})\, PI\left(\boldsymbol{o}, \min\{i \mid o_i \in \hat{T}\}\right)$$

exponential number of terms

- Dynamic programming algorithm

Algorithm 1 Computing $PD(\boldsymbol{o}, a)$

Input: prioritization game, prioritization $\boldsymbol{o}$, attack a

for $b = 0, 1, \ldots, B$ **do**

  $PD(\boldsymbol{o}, a, |T|, b) \leftarrow R_{a,o_{|T|}} \cdot F^*_{o_{|T|}}\left(\lfloor b / C_{o_{|T|}} \rfloor - 1\right)$

end for

for $i = |T| - 1, \ldots, 2, 1$ **do**

  for $b = 0, 1, \ldots, B$ **do**

    $PD(\boldsymbol{o}, a, i, b) \leftarrow R_{a,o_i} \cdot F^*_{o_i}(\lfloor b / C_{o_i} \rfloor - 1) + (1 - R_{a,o_i}) \sum_{j=0}^{\lfloor b / C_{o_i} \rfloor} \left[(F_{o_i}(j) - F_{o_i}(j-1)) \cdot PD(\boldsymbol{o}, a, b - j \cdot C_{o_i}, i+1) \right]$

  end for

end for

Return $PD(\boldsymbol{o}, a) := PD(\boldsymbol{o}, a, 1, B)$

Finding an Optimal Alert Prioritization Strategy

- Linear-programming based formulation
  - for each attack $a \in A$, solve

$$\max_{\boldsymbol{p}} \sum_{\boldsymbol{o} \in O} p_{\boldsymbol{o}} \cdot PD(\boldsymbol{o}, a)$$

subject to

$$\forall\, a' \in A: \quad \sum_{\boldsymbol{o} \in O} p_{\boldsymbol{o}} \cdot D(\boldsymbol{o}, a') \geq \Delta\left(K_{a'}\right)$$

where

$$D(o, a') = [(1 - PD(\boldsymbol{o}, a))G_a - (1 - PD(\boldsymbol{o}, a'))G_{a'}]$$

$$\Delta\left(K_{a'}\right) = K_a - K_{a'}$$

exponential number of possible orderings

  - output the solution that attains the lowest loss

Problem: Finding an improving column (i.e., ordering) is an **NP-hard** problem.

- Polynomial-time column generation approach

Algorithm 2 Greedy Column Generation

Input: prioritization game, reduced cost function $\bar{c}$

1: $\boldsymbol{o} \leftarrow \emptyset$
2: **while** $\exists t \in T \setminus \boldsymbol{o}$ **do**
3: $\quad \boldsymbol{o} \leftarrow \boldsymbol{o} + \mathrm{argmax}_{t \in T \setminus \boldsymbol{o}}\, \bar{c}(\boldsymbol{o} + t)$
4: **end while**
5: Return $\boldsymbol{o}$

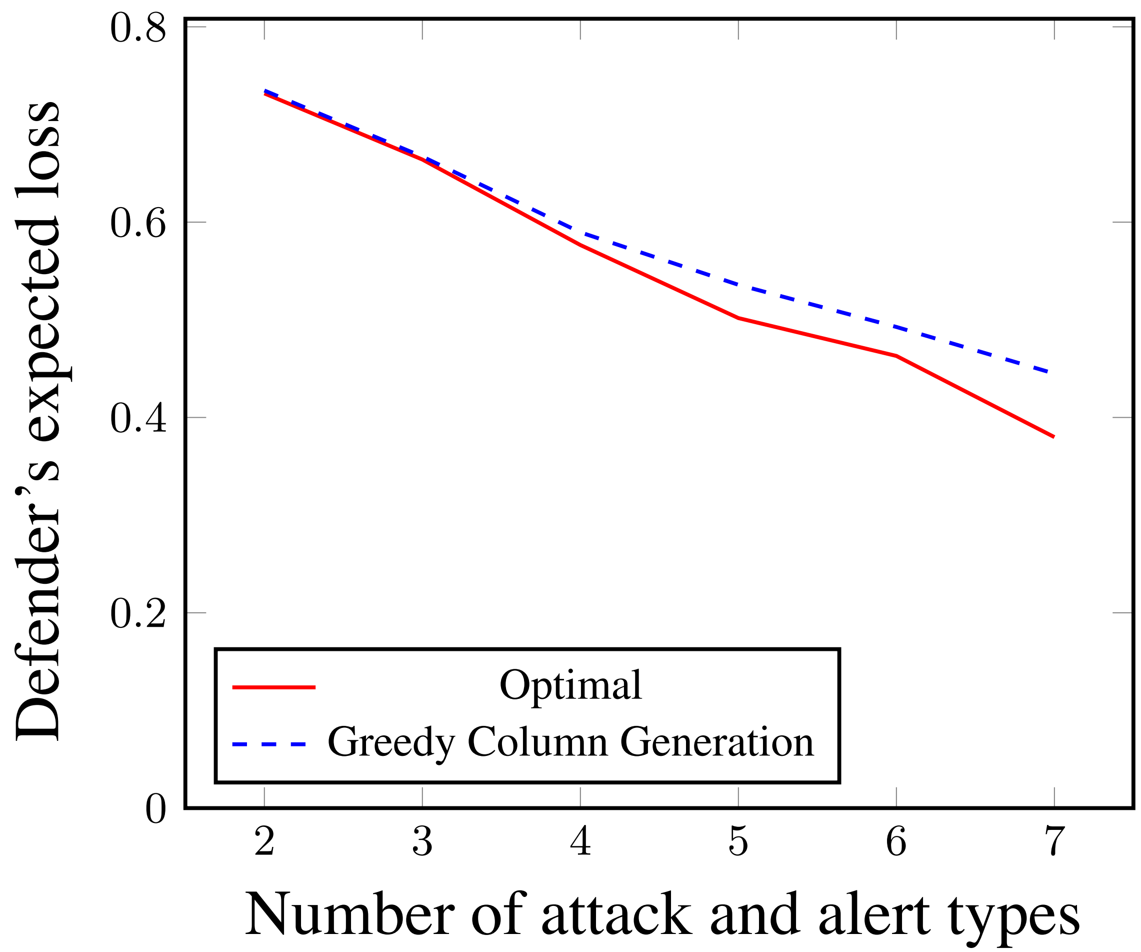
where

$$\bar{c}(\boldsymbol{o}) = PD(\boldsymbol{o}, a) + \sum_{a' \in A} y(\bar{O}, a') D(\boldsymbol{o}, a')$$

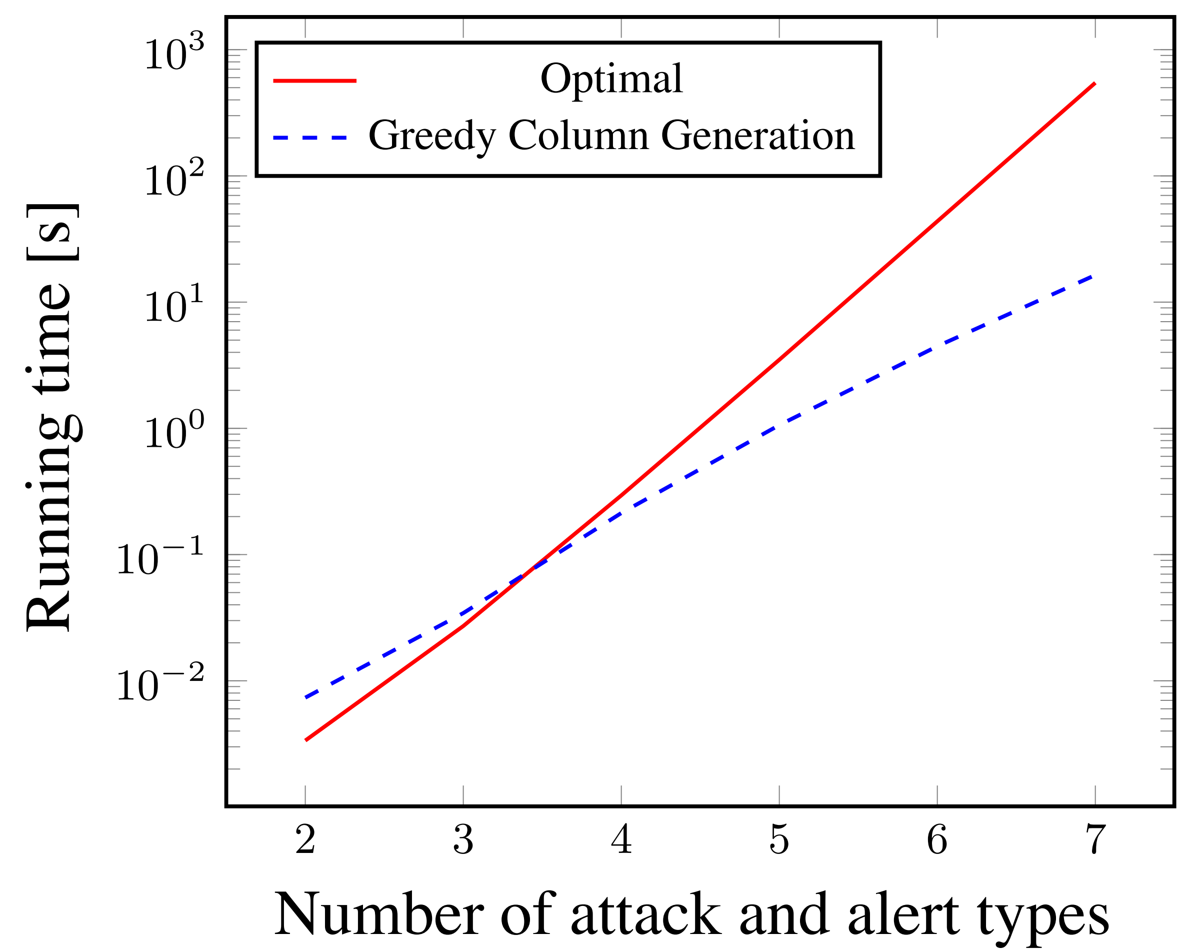
(i.e., reduced cost function)

Numerical Results - Synthetic Dataset

Defender's Loss



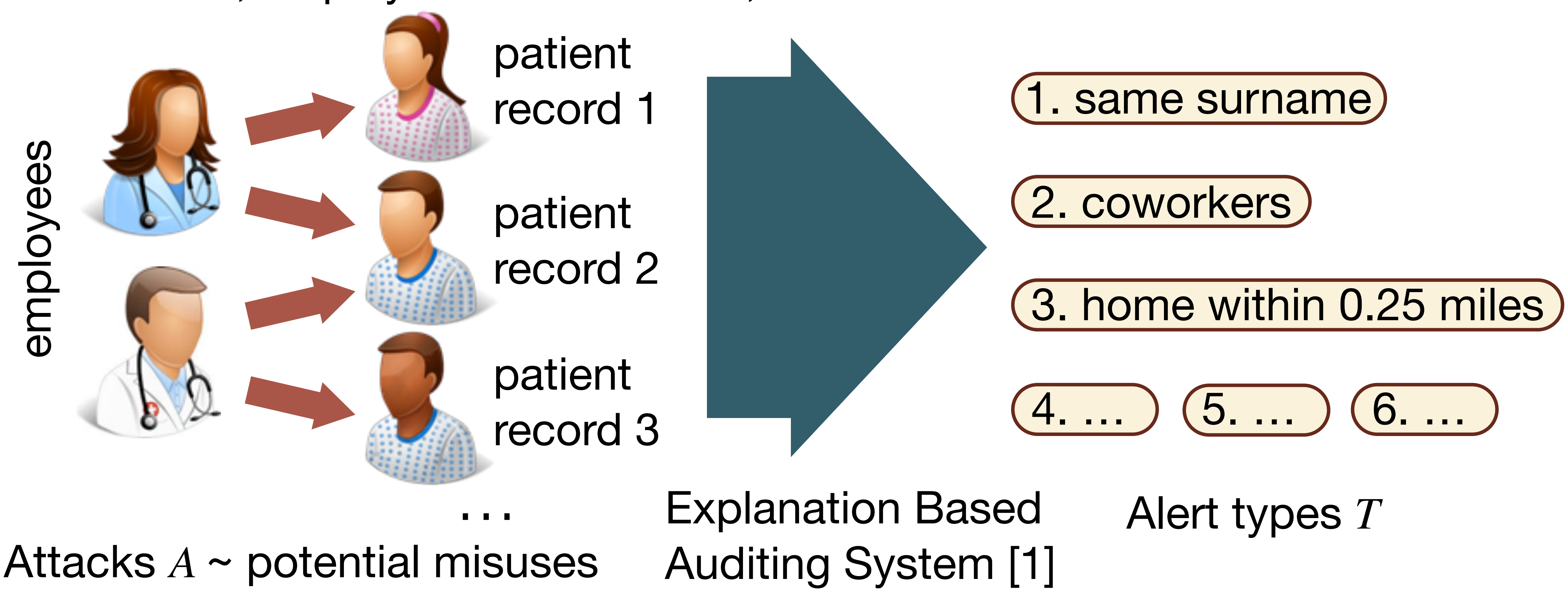
Running Time



$K_a = 0$, $C_t = 1$, $B = 5|T|$, D_a and G_a were drawn at random from $[0.5, 1]$, each $R_{a,t}$ is either 0 (with probability $1/3$) or drawn at random from $[0, 1]$, and every F_t has a Poisson distribution whose mean is drawn at random from $[5, 15]$.

Real-World Dataset: Electronic Medical Record System Alerts

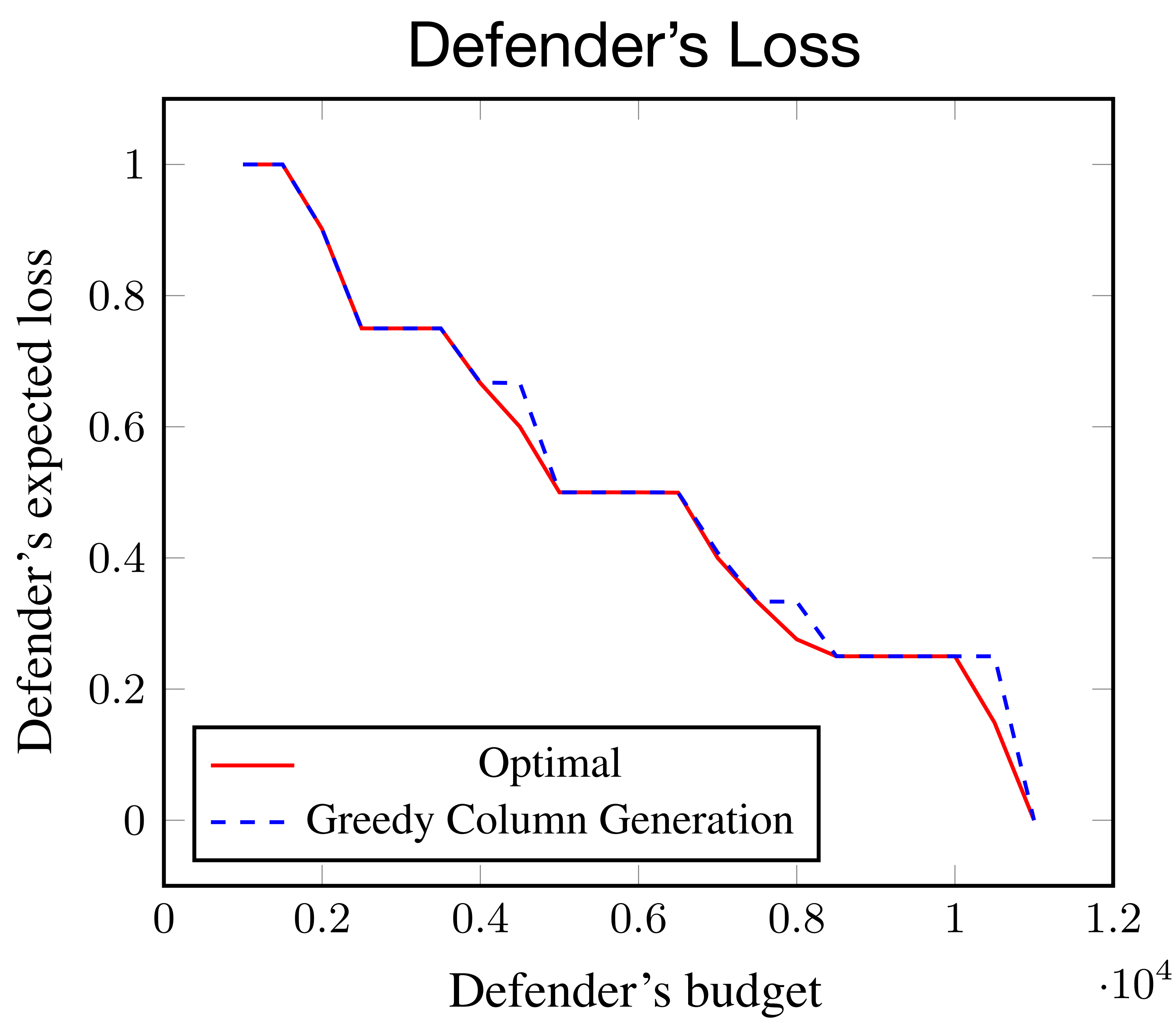
- Access logs from the **electronic medical record** (EMR) system in place at **Vanderbilt University Medical Center**
  - integrated with human-resources data to document medical department affiliation, employment information, and home addresses



[1] Fabbri, D., and LeFevre, K. 2013. Explaining accesses to electronic medical records using diagnosis information. Journal of the American Medical Informatics Association 20(1):52–60.

Numerical Results - Real-World Dataset

- Data collected from five consecutive weeks of access logs from 2016
- 8,481,767 accesses made by 14,531 users to 161,426 patient records, leading to a total of **863,989** alerts
- Approximated the distributions of false alerts using Poisson distributions
- In order to find optimal strategies, we restricted the alerts to 12 randomly selected patients



Conclusion & Future Work

- Prioritization of alerts is of crucial importance to the effectiveness of intrusion and misuse detection
- Result highlights
  - introduced **first model** of **alert prioritization** against **strategic adversaries**
  - showed that finding an optimal prioritization strategy is **NP-hard**
  - proposed an efficient **column-generation based approach**
  - evaluated numerically using **synthetic** and **real-world datasets**
- Future work
  - constant **approximation ratio** algorithms
  - modeling **multiple adversary types** as a Bayesian Stackelberg game

Thank you for your attention!

Questions?