Discovering Imperfectly Observable Adversarial Actions using Anomaly Detection*

Extended Abstract

Olga Petrova† Karel Durkota‡ Galina Alperovich†

Karel Horak† Michal Najman† Branislav Bosansky†‡

Viliam Lisy†‡

† Avast Software
{name.surname}@avast.com

‡ Dept. of Computer Science, FEE,
Czech Technical University in Prague
{name.surname}@fel.cvut.cz

ABSTRACT

Defenders in security problems often use anomaly detection (AD) to examine effects of (adversarial) actions and detect malicious behavior. Attackers seek to accomplish their goal (e.g., exfiltrate data) while avoiding the detection. Game theory can be used to reason about this interaction. While AD has been used in game-theoretic frameworks before, we extend the existing works to more realistic settings by (1) allowing players to have continuous action spaces and (2) assuming that the defender cannot perfectly observe the action of the attacker. We solve our model by (1) extending existing algorithms that discretize the action spaces and use linear programming and (2) by training a neural network using an algorithm based on exploitability descent, termed EDA. While both algorithms are applicable for low feature-space dimensions, EDA produces less exploitable strategies and scales to higher dimensions. In a data exfiltration scenario, EDA outperforms a range of classifiers when facing a targeted exploitative attacker.

KEYWORDS

anomaly detection, game theory, constrained learning

ACM Reference Format:
Olga Petrova, Karel Durkota, Galina Alperovich, Karel Horak, Michal Najman, Branislav Bosansky, and Viliam Lisy. 2020. Discovering Imperfectly Observable Adversarial Actions using Anomaly Detection. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 PROBLEM DESCRIPTION

Anomaly detection is used to detect malicious behavior in computer networks [8], fraud in financial transactions [1], or malicious behavior of software [5]. Attackers want to achieve some goal (e.g., exfiltrate data, commit fraud), but have their actions undetected. This can be modeled by integrating anomaly detection and game theory [3, 6, 7]. The existing works assume discrete features or perfect observability of the effects of attacker's actions by the defender. It limits their usability in realistic settings, e.g., if an attacker's network traffic is mixed with the traffic of benign users, or transactions misusing a credit card mix with its regular usage.

In our two-player game, the defender is a stochastic classifier which observes events—points in n-dimensional feature space—and estimates a probability of inspecting the event for maliciousness. The attacker chooses an action that generates an event in the same feature space. If the event gets undetected, the attacker receives a corresponding reward.

Generalized Classification Game is a tuple $G = (\mathcal{F}, C, R, P_D, \phi, \mathcal{T})$ where $\mathcal{F}$ is the n-dimensional real-valued feature space, each feature f^i is bounded by $[L^i, U^i]$, $i = 1, \ldots, n$. C is a set of all classifiers of the defender and $c \in C$ is a function $c : \mathcal{F} \to [0, 1]$, $c(f)$ is the probability that an event $f \in \mathcal{F}$ gets inspected. An uninspected attacker's action $f_a \in \mathcal{F}$ yields a non-negative reward $R(f_a) \geq 0$ for the attacker. $\phi \in [0, 1]$ is the maximal allowed false-positive rate (FPR) of the classifier. We assume that the benign events are samples from a distribution P_D, hence the expected FPR of a classifier c, denoted $\Phi_D(c)$, satisfies $\Phi_D(c) = \mathbb{E}_{f \sim P_D}\, c(f)$. Finally, $\mathcal{T}$ corresponds to an observation transformation function: the defender observes feature vectors f_o sampled from a probability distribution $\mathcal{T}(f_a) \in \Delta(\mathcal{F})$ for attacker's action $f_a \in \mathcal{F}$. There are no restrictions on function $\mathcal{T}$ but we focus on two specific cases: an identity, and an additive combination of effects of attacker's and benign actions $f_b \sim P_D$, i.e. $\mathcal{T}(f_a) = f_a + P_D$, i.e., $\Pr_{\mathcal{T}}[f = f_a + f_b \mid f_a] = \Pr_{P_D}[f_b]$.

Attacker's utility in the game corresponds to $u_a(c, f_a) = (1 - \rho_c(f_a)) \cdot R(f_a)$ where $\rho_c(f_a) = \mathbb{E}_{f \sim \mathcal{T}(f_a)}[c(f)]$ is the expected probability that c classifies action f_a (based on observations $f \sim \mathcal{T}(f_a)$) as anomalous. We assume the zero-sum game, hence the attacker maximizes the value and the defender minimizes. The solution corresponds to a maximin (or a Stackelberg equilibrium). The defender seeks a classifier c^* that minimizes the attacker's utility under a FPR constraint *assuming* that the attacker plays a best response:

$$c^* = \arg\min_{c \in C} \{ \max_{f_a \in \mathcal{F}} u_a(c, f_a) \} \quad \text{s.t. } \Phi_{P_D}(c) \leq \phi \tag{1}$$

We use $BR_a(c)$ to denote the best response (BR) of the attacker to the classifier c, where $BR_a(c) = \arg\max_{f_a} u_a(c, f_a)$.

*This work was partially supported by the OP VVV MEYS funded project CZ.02.1.01/0.0/0.0/16 019/0000765 "Research Center for Informatics"

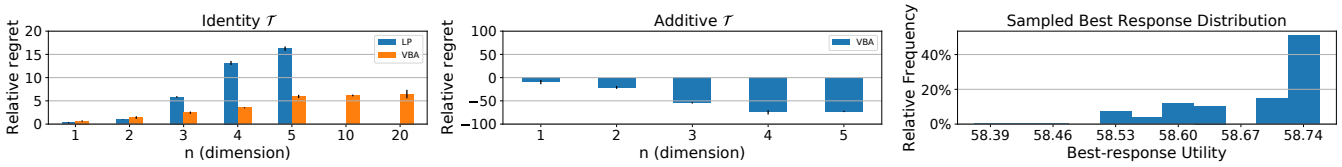

Figure 1: (left) Relative regret of LP and EDA compared to the optimal game-values in case of identity transformation function $\mathcal{T}$; (middle) Relative regret of EDA compared to LP for general transformation function $\mathcal{T}$; (right) Evaluation of variance of sampled best response with $30,000$ samples used for data exfiltration experiments.

Algorithm 1: Exploitability Descent for Adversarial Anomaly Detection (EDA).

$\theta^0 \leftarrow$ initial random NN, $\lambda^0 \leftarrow 1.0$, $i \leftarrow 0$
while *termination condition is not met* **do**
  $f_a \leftarrow BR_a(c^i)$; $\hat{\phi}^i \leftarrow \Phi_D(c^i)$
  $\mathcal{L}^i \leftarrow -\mathbb{E}_{f\sim\mathcal{T}(f_a)}[\log(c^i(f))] + \lambda^i(\hat{\phi}^i - \phi)$
  $\theta^{i+1} \leftarrow \theta^i - \alpha\nabla_\theta\mathcal{L}^i$; $\lambda^{i+1} \leftarrow \max\{0, \lambda^i + \beta\nabla_\lambda\mathcal{L}^i\}$
  $i \leftarrow i + 1$
$\theta \leftarrow OutputClassifier(\{\theta^0, \theta^1, \dots\}, \phi)$
Output: θ

2 ALGORITHMS

Linear Programming (LP) has been used for solving simpler models [7]. Below, we propose a more general formulation. We uniformly discretize the feature space $\mathcal{F}$ into d^n grid cells $\bar{f} \subset \mathcal{F}$ (each dimension with d bins), creating the set $\mathcal{F}_d$ containing all grid cells. Now both players have a finite number of actions – attacker chooses a grid cell $\bar{f}_a$ (assuming attacker chooses the action that maximizes the reward function within each grid cell), the defender chooses a probability of inspection of each grid cell (variables $c(\bar{f}_o)$). Let $\Pr[\bar{f}|A]$ denote the probability density in a cell $\bar{f}$ given the probability distribution A, the LP is:

$$\min_{\{c(\bar{f}_o):\bar{f}_o\in\mathcal{F}_d\}} V \tag{2}$$

$$\text{s.t.:}(\forall\bar{f}_a \in \mathcal{F}_d) : R(\bar{f}_a) \sum_{\bar{f}_o\in\mathcal{F}_d} \Pr[\bar{f}_o|\mathcal{T}(\bar{f}_a)](1 - c(\bar{f}_o)) \le V \tag{3}$$

$$\sum_{\bar{f}_o\in\mathcal{F}_d} \Pr[\bar{f}_o|P_D]c(\bar{f}_o) \le \phi \tag{4}$$

$$(\forall\bar{f}_o \in \mathcal{F}_d) : 0 \le c(\bar{f}_o) \le 1, \tag{5}$$

where V is the value of the game, Eqs. (3) are best-response constraints, and Eq. (4) ensures maximal FPR of ϕ. To overcome overfitting to a data sample D, we approximate the distribution P_D from D using Kernel Density Estimation (KDE) with kernel bandwidth parameter h (selected using a binary search). The approximated distribution is then used in (4).

Exploitability Descent for Adversarial Anomaly Detection (EDA) (see Algorithm 1) is based on *exploitability descent* [9, 11]. EDA iteratively solves the problem in Eq. (1) – in each iteration, EDA first estimates the BR of the attacker (f_a) to the current classifier (c) and, second, it updates the classifier c with stochastic gradient descent w.r.t. the FPR constraint ϕ. We model the classifier with a neural net c_θ, parametrized by weights θ.

To overcome the problem of vanishing gradients, we construct an upper bound of the outer minimization problem of (1) by taking the logarithm of the criterion and minimize the upper bound instead ($-\mathbb{E}_{f\sim\mathcal{T}(f_a)}[\log(c_\theta(f))]$). We use Lagrangian ($\mathcal{L}$) relaxation procedure [2, 4] to move the FPR constraint into the objective to obtain unconstrained problem as follows:

$$\max_{\lambda\ge 0}\min_{\theta\in\Theta} -\mathbb{E}_{f\sim\mathcal{T}(f_a)}[\log(c_\theta(f))] + \lambda\cdot(\Phi_D(c_\theta) - \phi) \tag{6}$$

EDA stores all pareto-optimal classifiers in the space of the FPR and the attack value. Therefore, final classifier is selected as a convex combination such that the value is minimal for the desired FPR ϕ.

3 EXPERIMENTAL VALIDATION

In our experiments, EDA uses an NN with 3 fully-connected layers with $32 + 2n$, $32 + 2n$, $16 + 2n$, and 1 (output) neurons (n is the dimension of the feature space). Sigmoid function is used for the last neuron, ReLU activation function is used for all other neurons. For the comparison, we need to approximate BR of the attacker. We use random sampling with hill-climbing in EDA, but we use random sampling without hill-climbing for the algorithm comparison due to discretization in LP. Figure 1 (right) shows a histogram of values of $10,000$ repeated estimations based on sampled BR with $30,000$ samples used for real-world experiments demonstrating a low variance and stability of this approximation.

Figure 1 (left) and (middle) show that EDA achieves smaller regret (i.e. more robust strategies) than LP for both transformation function on synthetic data and that EDA is able to scale to higher dimensions than LP.

With the real-world data, we compare against standard anomaly detectors. We focus on the data exfiltration case via DNS protocol and use 20,000 anonymized real-world DNS queries. We consider three features – the sum of lengths of all queries in a group within the time interval, the sum of entropy, and the number of special symbols. Reward function of the attacker is a multiplication of values of the first two features, corresponding to the amount of leaked information. We compare EDA with Principal Component Analysis (PCA), Isolation Forest, K-nearest neighbors with largest distances to the k-th neighbor as the outlier score and cluster-based local outlier factor algorithm implemented in [10]. In our experiments, EDA found the least exploitable strategy (attacker's best-response value 58.75), while the best traditional anomaly detector, PCA, has higher exploitability (attacker's best-response value 85.83).

We have shown a method EDA, not requiring discretization, that beats PCA and other baselines in exploitability while maintaining scalability on higher dimensional problems.

REFERENCES

[1] Mohiuddin Ahmed, Abdun Naser Mahmood, and Md Rafiqul Islam. 2016. A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems* 55 (2016), 278–288.

[2] D.P. Bertsekas. 1999. *Nonlinear programming*. Athena Scientific.

[3] Michael Brückner, Christian Kanzow, and Tobias Scheffer. 2012. Static prediction games for adversarial learning problems. *Journal of Machine Learning Research* 13, Sep (2012), 2617–2654.

[4] Yinlam Chow, Mohammad Ghavamzadeh, Lucas Janson, and Marco Pavone. 2017. Risk-constrained reinforcement learning with percentile risk criteria. *The Journal of Machine Learning Research* 18, 1 (2017), 6070–6120.

[5] Gianluca Dini, Fabio Martinelli, Andrea Saracino, and Daniele Sgandurra. 2012. MADAM: a multi-level anomaly detector for android malware. In *International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security*. Springer, 240–253.

[6] Lemonia Dritsoula, Patrick Loiseau, and John Musacchio. 2017. A game-theoretic analysis of adversarial classification. *IEEE Transactions on Information Forensics and Security* 12, 12 (2017), 3094–3109.

[7] Karel Durkota, Viliam Lisý, Christopher Kiekintveld, Karel Horák, Branislav Bošanský, and Tomáš Pevný. 2017. Optimal strategies for detecting data exfiltration by internal and external attackers. In *International Conference on Decision and Game Theory for Security*. Springer, 171–192.

[8] Pedro Garcia-Teodoro, Jesus Diaz-Verdejo, Gabriel Maciá-Fernández, and Enrique Vázquez. 2009. Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security* 28, 1-2 (2009), 18–28.

[9] Edward Lockhart, Marc Lanctot, Julien Pérolat, Jean-Baptiste Lespiau, Dustin Morrill, Finbarr Timbers, and Karl Tuyls. 2019. Computing Approximate Equilibria in Sequential Adversarial Games by Exploitability Descent. *arXiv preprint arXiv:1903.05614* (2019).

[10] Yue Zhao, Zain Nasrullah, and Zheng Li. 2019. PyOD: A Python Toolbox for Scalable Outlier Detection. *Journal of Machine Learning Research* 20, 96 (2019), 1–7. http://jmlr.org/papers/v20/19-011.html

[11] Najman, M. 2019. Adversarial Machine Learning for Detecting Malicious Behavior in Network Security. Master's thesis, FEE, Czech Technical University in Prague.