RESONANT: Reinforcement Learning-based Moving Target Defense for Credit Card Fraud Detection

George Abdel Messih

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science and Application

Peter Beling, Chair
Jin-Hee Cho, Co-Chair
Tyler Cody
Bo Ji

November 10, 2023
Blacksburg, Virginia

Keywords: Iterated games, adversarial learning, moving target defense, credit card, fraud detection, reinforcement learning, machine learning.

RESONANT: Reinforcement Learning-based Moving Target Defense for Credit Card Fraud Detection

George Abdel Messih

(ABSTRACT)

According to *security.org*, as of 2023, 65% of credit card (CC) users in the US have been subjected to fraud at some point in their lives, which equates to about 151 million Americans. The proliferation of advanced machine learning (ML) algorithms has also contributed to detecting credit card fraud (CCF). However, using a single or static ML-based defense model against a constantly evolving adversary takes its structural advantage, which enables the adversary to reverse engineer the defense's strategy over the rounds of an iterated game. This paper proposes an adaptive *moving target defense* (MTD) approach based on *deep reinforcement learning* (DRL), termed `RESONANT`, to identify the optimal switching points to another ML classifier for credit card fraud detection. It identifies optimal moments to strategically switch between different ML-based defense models (i.e., classifiers) to invalidate any adversarial progress and always stay a step ahead of the adversary. We take this approach in an iterated game theoretic manner where the adversary and defender take turns to take their action in the CCF detection contexts. Via extensive simulation experiments, we investigate the performance of our proposed `RESONANT` against that of the existing state-of-the-art counterparts in terms of the mean and variance of detection accuracy and attack success ratio to measure the defensive performance. Our results demonstrate the superiority of `RESONANT` over other counterparts, including static and naïve ML and MTD selecting a defense model at random (i.e., Random-MTD). Via extensive simulation experiments, our results show that our proposed `RESONANT` can outperform the existing counterparts up to

two times better performance in detection accuracy using AUC (i.e., Area Under the Curve of the Receiver Operating Characteristic (ROC) curve) and system security against attacks using attack success ratio (ASR).

RESONANT: Reinforcement Learning-based Moving Target Defense for Credit Card Fraud Detection

George Abdel Messih

(GENERAL AUDIENCE ABSTRACT)

According to *security.org*, as of 2023, 65% of credit card (CC) users in the US have been subjected to fraud at some point in their lives, which equates to about 151 million Americans. The proliferation of advanced machine learning (ML) algorithms has also contributed to detecting credit card fraud (CCF). However, using a single or static ML-based defense model against a constantly evolving adversary takes its structural advantage, which enables the adversary to reverse engineer the defense's strategy over the rounds of an iterated game. This paper proposes an adaptive defense approach based on *artificial intelligence* (AI), termed `RESONANT`, to identify the optimal switching points to another ML classifiers for credit card fraud detection. It identifies optimal moments to strategically switch between different ML-based defense models (i.e., classifiers) to invalidate any adversarial progress and always stay a step ahead of the adversary. We take this approach in an iterated game theoretic manner where the adversary and defender take turns to take their action in the CCF detection contexts. Via extensive simulation experiments, we investigate the performance of our proposed `RESONANT` against that of the existing state-of-the-art counterparts in terms of the mean and variance of detection accuracy and attack success ratio to measure the defensive performance. Our results demonstrate the superiority of `RESONANT` over other counterparts, showing that our proposed `RESONANT` can outperform the existing counterparts by up to two times better performance in detection accuracy and system security against attacks.

Acknowledgments

This work is partly supported by the National Securittates Government under Grant Number H98230-21-1-0322 and the US Army Research Office with a grant W911NF-20-2-0140. This work is authorized to reproduce and distribute reprints notwithstanding any copyright notation herein.

RESONANT: Reinforcement Learning-based Moving Target Defense for Credit Card Fraud Detection

George Abdel Messih*†, Tyler Cody†, Peter Beling†‡, Jin-Hee Cho*‡
*Department of Computer Science, Virginia Tech, Falls Church, VA 22043 USA
†National Security Institute, Virginia Tech, Arlington, VA 22203 USA
‡Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061 USA

***Abstract*—According to *security.org*, as of 2023, 65% of credit card (CC) users in the US have been subjected to fraud at some point in their lives, which equates to about 151 million Americans. The proliferation of advanced machine learning (ML) algorithms has also contributed to detecting credit card fraud (CCF). However, using a single or static ML-based defense model against a constantly evolving adversary takes its structural advantage, which enables the adversary to reverse engineer the defense's strategy over the rounds of an iterated game. This paper proposes an adaptive *moving target defense* (MTD) approach based on *deep reinforcement learning* (DRL), termed `RESONANT`, to identify the optimal switching points to another ML classifier for credit card fraud detection. It identifies optimal moments to strategically switch between different ML-based defense models (i.e., classifiers) to invalidate any adversarial progress and always stay a step ahead of the adversary. We take this approach in an iterated game theoretic manner where the adversary and defender take turns to take their action in the CCF detection contexts. Via extensive simulation experiments, we investigate the performance of our proposed `RESONANT` against that of the existing state-of-the-art counterparts in terms of the mean and variance of detection accuracy and attack success ratio to measure the defensive performance. Our results demonstrate the superiority of `RESONANT` over other counterparts, including static and naïve ML and MTD selecting a defense model at random (i.e., Random-MTD). Via extensive simulation experiments, our results show that our proposed `RESONANT` can outperform the existing counterparts up to two times better performance in detection accuracy using AUC (i.e., Area Under the Curve of the Receiver Operating Characteristic (ROC) curve) and system security against attacks using attack success ratio (ASR).**

***Index Terms*—Iterated games, adversarial learning, moving target defense, fraud detection, reinforcement learning.**

I. Introduction

A. Motivation

Advances in modern technologies have introduced rapid growth of e-commerce and digital marketplaces, which significantly increase the convenience of online shopping with fast and secure digital transactions. However, these digital transactions are still susceptible to fraud. Credit card fraud (CCF) is a widespread business problem worth billions of dollars across the world [1, 2]. In practice, many artificial intelligence and machine learning (AI/ML) models have been used to defend against fraudulent transactions [1–3]. Nevertheless, fraudsters have continuously developed new techniques to infiltrate these online-based systems with malicious intent, such as evasion, poisoning, and exploratory attacks [2, 3]. Most existing approaches aim to develop stronger yet static defensive ML models that often fail because they cannot deal with advanced adversaries capable of reverse-engineering the defender's model over time [4, 5]. Thus, the need for novel, effective, dynamic ML defenses adaptable to the ever-evolving adversarial attack strategies has been recognized [6].

There are **drawbacks to existing defense strategies**, such as adversarial training (AT) and generative adversarial networks (GANs). For instance, AT assumes that the fraudulent transactions the defense trains on will be very similar to the fraud faced in the real world [6]. In addition, the min-max nature of GANs makes similar assumptions regarding the knowledge of the adversary's pay-off structure [7, 8]. Since the real-world adversary is unknown and evolving, these assumptions, even if accurate initially, often lead to performance deterioration over time when the defense is deployed in real-world settings. Adversaries may not be rational, may vary their goals and attack strategies, and may directly exploit such assumptions. Even when the assumptions are appropriate, Arora et al. [8] showed how imposing realistic system conditions on the discriminator, such as finite capacity and the limited number of training steps, can disprove the GANs' ability to reach an equilibrium.

B. Research Goal & Key Ideas

We aim to avoid such assumptions by introducing an adaptable ML/AI-based defense. We adopt the principles of *moving target defense* (MTD), a proactive defense technique to increase attackers' uncertainty by changing attack surfaces. This approach will challenge the attacker to predict the defender's actions and adapt to their defenses [9]. Nonetheless, the MTD's effectiveness is significantly influenced by the defender's ability to predict potential attacks or how much knowledge the adversary has about the defense model [9, 10]. To address this limitation, we leverage the *reinforcement learning* (RL) technique to intelligently determine what ML classifier to ensure the robustness of the fraud detection process against CCF attacks. The RL agent chooses an ML classifier to maximize its reward per round. The RL agent will allow learning via trial and error and thus does not need the adversary's knowledge due to the nature of RL's

autonomous decision-making process. We name our proposed approach RESONANT, representing 'Reinforcement lEarning-baSed mOviNg tArget defeNse deTection' for CCF.

C. Key Contributions

Our proposed RESONANT has the **key contributions**:

1) RESONANT is a novel MTD approach that continuously changes an ML classifier to detect CCF. The underlying principle is to eliminate the nature of a static ML approach, whose knowledge and security vulnerabilities can be easily exploited by adversaries.
2) RESONANT leverages RL to create a model-free defense without the strong assumptions towards adversarial behaviors, often generating modeling errors and performance deterioration. In addition, for more realistic modeling of attack-defense interactions, we consider their iterative interactions in the context of CCF.
3) Via extensive simulation experiments, we demonstrate the outperformance of the proposed RESONANT over the considered counterparts in terms of the detection effectiveness with minimum variance across the iterated interactions, attack success ratio, and algorithmic efficiency.

D. Paper Structure

This paper is structured as follows. Section II discusses the overview of the related work, mainly in terms of reinforcement learning-based moving target defense and ML-based CCF detection. Section III discusses the background information of this domain, i.e. defining possible threat models in CCF and existing ML optimization methods for ML robustness. Section IV describes the network and threat models considered in this work. Section V describes the details of the proposed RESONANT in realizing MTD using RL. Section VI describes datasets, comparing defense mechanisms and metrics to conduct valid experimental environments. Section VII demonstrates the experimental results and analyzes their overall trends along with the underlying reasons. Section VIII concludes the paper and suggests future work directions.

II. Related Work

This section discusses the overview of the state-of-the-art RL-based MTD for cybersecurity applications and ML-based CFF mechanisms.

A. RL-based MTD for Cybesecurity

Eghtesad et al. [11] proposed an RL-based MTD strategy applied in the cyber-security domain, where the RL agent's action set is either to do nothing or to select a server from the network and re-image it to reset the adversarial infiltration of that server. Eghtesad et al. [12] further expanded on Eghtesad et al. [11] by proposing and assessing various extensions to their previously used RL-based defense, such as dueling and double Q-Network agents. Chowdhary et al. [10] proposed a similar RL-MTD framework in the cybersecurity domain where the RL agent has three possible actions [no action, network shuffle, IP mutation]. This work showed that utilizing an RL-based MTD against an adaptive attacker produces better results than deploying a random selection-based MTD. Zhu et al. [13] also proposed an iterative RL-based MTD for the cybersecurity domain. The authors mathematically proved that their RL algorithm can perfectly reach convergence against another RL algorithm simulating the attacker's actions. Yet, the authors designed their framework in a min-max nature between the two RL agents and evaluated the performance of their proposed defense accordingly. Their findings resulted in the same vulnerabilities of other min-max frameworks, such as GANs. Yoon et al. [14] proposed a multi-agent deep RL (DRL)-based MTD approach to enhance an in-vehicle network's performance by making two decisions related to link bandwidth allocation to meet quality-of-service (QoS) requirements and the frequency of triggering IP shuffling as an MTD technique. Chai et al. [15] proposed a DRL algorithm to decide the optimal duration between subsequent network shuffles for the security of cyber-physical systems to minimize defense resource consumption while maintaining a secure integration of wearable devices in people's everyday lives. Kim et al. [16] proposed a DRL-based automated cybersecurity framework to run scalable and effective network traffic inspection for threat detection and network address shuffling as an MTD operation to proactively handle attacks.

Limitations: However, unlike the above works discussed, our work is the first to propose a model-free RL-based MTD in the ML robustness domain against attackers in the context of CCF detection. No prior work has applied the concept of MTD to increase the diversity of ML classifiers for CCF, leading to effective changes in the attack surface.

B. ML-based Credit Card Fraud Detection

Both Vimal et al. [1] and Zhinin-Vera et al. [17] proposed RL agents that mimic a typical ML classifier but with more advanced training metrics and learning models. Each RL agent described in these efforts interacts with a single transaction at a time and formulates its state from that transaction's features. The agent then takes a binary classification action of 0 for non-fraud or 1 for fraud. Dornadula and Geetha [18] proposed an ML-based CCF detection mechanism where customers are clustered into groups based on their historical transactions and behavioral patterns. The authors demonstrated the conventional ML-based CCF detection pipeline, where the imbalanced data was addressed using SMOTE. They used the conventional ML classifiers, such as decision trees, for CCF detection. Khatri et al. [19], Awoyemi et al. [20] and Dhankhad et al. [21] demonstrated comparative analysis frameworks on the effectiveness of the most common supervised learning ML models when applied in the CCF detection domain, such as k-nearest neighbors (KNN), Logistic Regression and Random Forest. The authors also provided a basic framework to address the class imbalance in CCF datasets, using either over-sampling, under-sampling, or both.

Limitations: However, both RL models in [1, 17] are only capable of making binary classification decisions, such as fraud or non-fraud. On the other hand, our proposed

RESONANT is designed for an RL agent to make decisions at the meta-level of a game between the defender and the adversary to determine the optimal classifier to utilize at each step of the game. This achieves effective, reliable MTD strategies to maximize the system performance in terms of high detection accuracy and low attack success ratio. Further, unlike the static ML-based CCF detection methods in [18–21], our proposed RESONANT incorporates RL to develop an effective MTD strategy to realize a dynamic, proactive defense capable of counteracting sophisticated attacks in CCF detection.

III. Background

Currently, most ML robustness research in the CC fraud detection domain is focused on one of two major areas, with many variations within each. Firstly, defining the existing threat models to have a better understanding of adversarial capabilities. Secondly, optimizing classifiers for better adversarial detection, like in Generative Adversarial Networks (GANs), or adversarial training.

A. Threat Models

[22] and [23] Define the following CC fraud types:

- Bankruptcy CC fraud: where a user applies and uses a credit card while being in a state of bankruptcy. Hence, the bank will be unable to collect the loaned amount to that user and will incur that loaned amount as a loss.
- Application CC fraud: where a fraudster uses false information to apply for a credit card.
- Behavioural CC fraud:
  - Theft/ offline CC fraud: where the fraudster uses a physically stolen card to conduct illegal transactions.
  - Counterfeit/ online CC fraud: where the fraudster was able to attain a credit card's information without keeping the actual card, and then uses that information to make illegal online transactions.

Furthermore, [24] and [25] provide the following formulation of adversarial threat models:

- Evasion Attacks: Evasion attacks are the most common type of adversarial attacks, where the adversary provides samples for the defender's classifier to evaluate with the intention of increasing the misclassification rate. The adversary adjusts the malicious data, such that it evades the defender system during the testing phase.
- Poison Attacks: Poison Attacks target the defense system by injecting carefully designed samples into the training data of the defender's ML model to compromise its learning process. For instance, generative adversarial networks (GANs) are an example of poisoning attacks, where a generative model G attempts to generate data that is undetectable to the defender/ discriminator model [26]. By contaminating the training process, the adversary compromises the whole system.
- Exploratory Attacks: Exploratory attacks rely on exploiting the underlying classification model and defense architecture, where the adversary sends numerous inquiries to the defense system to gain as much knowledge as possible about the defender's learning algorithm. Each attack's effectiveness is dependent on the adversary's knowledge of the defender.

B. Classifier Optimization

- GANs: GANs is a framework, in which two models are simultaneously trained in an iterative minimax two-player game: a generative model G that captures the data distribution and generates new fraudulent samples in each round of the game, and a discriminative model D that classifies if a sample came from the training data or G. The goal for G is to maximize the probability of D making a mistake, while D trains to minimize that same probability [26]. The overall aim of this framework is that by pitting these two models against each in this iterative simulation both models are actually getting more efficient at each of their tasks [7]. Thus, GANs are often used to produce an optimized classifier able to accurately detect and reject fraudulent samples.
  However, the minimax nature of GANs assumes knowledge of the adversary's pay-off/ utility function and assumes a fixed and rational adversary [7]. Additionally, there is research showing that imposing realistic parameters on the discriminator defined in GANs, like finite capacity and limited number of training steps could disprove that the GANs framework could reach an equilibrium [4] [7].
- Adversarial Training: Similarly, adversarial training aims to optimize classifiers' ability to detect fraud by injecting labeled fraudulent samples in the training process of the classifier. The goal is that the model is able to learn the data distribution of the fraudulent data, leading to a more accurate adversarial detection process. Accordingly, this method can act as an effective first line of defense.
  However, in order for this method to be effective against strong attacks the model needs to have been previously trained on such attacks, which may not be the case when the model is implemented in the real world [27] and [28]. It also assumes that the data distribution that the defense trains on will be identical to the data distribution of the fraud faced in the real world [6].

IV. System Model

This section describes the considered system model in terms of network and threat models.

A. Network Model

As shown in Fig. 1, we consider an enterprise network providing credit card (CC) services. In this network, a cloud server exists to process all online CC transactions and customers' deposits or withdrawals. Simultaneously, fraudsters send in fraudulent transactions to that cloud server, aiming to make illegal CC transactions.

B. Threat Model

We concern CCF adversaries aiming to increase their fraud infiltration rate by passing multiple fraud transactions. The

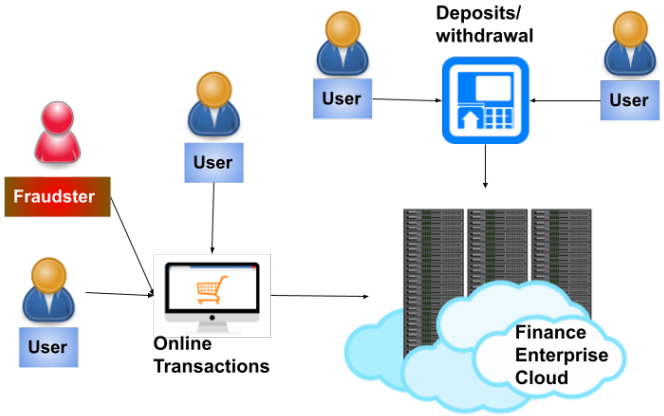


Fig. 1: Example of a credit card enterprise network model.

adversaries observe what transactions are accepted and learn the infiltration patterns by repeating this process. By learning the system's vulnerabilities, the adversaries will make more intelligent fraudulent transactions to achieve the highest fraud acceptance rate. The most common adversarial attack type in CCF is *evasion attacks* by injecting fraudulent or adversarial transactions into the defender's system to maximize misclassification. The adversary designs and manipulates the malicious transactions to evade the defense system during the testing phase [25].

V. Proposed Approach: RESONANT

This section describes the overview of the proposed RESONANT in terms of attack-defense interactions, resolving class imbalance, and the RL agent's MTD decision-making.

A. Attack-Defense Interactions

Inspired by the game-theoretic framework in [29], we create a 50-round iterative game where both the adversary and the defender aim to maximize their respective utilities, simulating the attack-defense iterative interactions in the CCF domain. The attacker aims to maximize the number of accepted fraud transactions by selecting which transactions to over-sample and inject into the defender's system in each round. On the other hand, the defender aims to detect fraud transactions based on detection accuracy metrics (e.g., ROC-AUC score). In the conventional setting, the defender can achieve this through two potential actions: retraining its ML classifier between rounds or implementing a random-MTD strategy. In our work, we introduce a defender agent that can autonomously make an optimal MTD decision based on RL.

Furthermore, we implement a label delay window (LDW) between undetected fraudulent transactions and receiving their truth labels, representing the delay window between undetected fraud occurrences and customers reporting these frauds.

B. Attacker's and Defender's Goals and Actions

1) **Defender's Goal and Action:** The RL-MTD defense agent aims to maximize its reward function. It achieves its goal by choosing an optimal action, an ML classification model to deploy against CCF adversaries based on the environment's current state to maximize its reward.

2) **Attacker's Goal and Action:** We model adversarial behavior by creating an artificial fraud generation method performing evasion attacks under varying defense strategies controlled by our RESONANT in the iterative games modeled in this work. The fraud generation method was created using *the Synthetic Minority Oversampling Technique (SMOTE)-based Offensive Policy* [30], a semi-black-box targeted evasion attack. We choose SMOTE because of its merit in efficiency [31]. In this method, the adversary collects the accepted frauds from each round and applies the SMOTE to generate new fraudulent samples, which fall within the same decision space previously accepted by the defender's classifier. Next, the adversary injects the newly generated fraudulent transactions into the next round's defender test set. This process is repeated each round to generate new fraudulent samples. The adversary's goal in each round is to maximize the number of accepted fraudulent transactions, i.e., maximizing false negatives. Accordingly, in each round, the adversary must select which credit card transactions to over-sample using SMOTE and inject it into the next round's test data.

In addition, we demonstrate that the proposed testing method can highlight defense vulnerabilities that may not have shown in the single-round test typically used by other research efforts [19–21]. For instance, the static ML defense experiment shows how the defense performs well in the first round while the adversary is able to infiltrate the defender's system as more interactions occur in the iterative game. Accordingly, this proves that the snapshot type of experiments typically run in most other research efforts in the CCF detection domain is insufficient to evaluate the defense model's effectiveness.

C. Resolving Class Imbalance

One of the well-known challenges in most CCF detection models is a class imbalance issue. There are many more cases of legitimate transactions than fraudulent transactions. For instance, our data has only $\sim$0.1% fraud. Accordingly, if the ML model training is performed on the unbalanced data, the model will learn to classify everything as non-fraud and have a 99.9% accuracy. However, none of the fraud will be detected. Therefore, data distribution between the fraudulent minority class and the non-fraudulent majority class should be balanced as the first step. There are multiple approaches to deal with imbalanced data, such as oversampling the minority class or under-sampling the majority class [32]. Over-sampling possibly causes over-fitting, while under-sampling leads to losing valuable information on the majority classes [33].

To avoid losing any useful information from under-sampling as well as provide more data for training a DRL model (e.g., DQN) of the RL-MTD agent, we use SMOTE to generate additional fraudulent transactions, the minority class [34]. While there are more advanced options for data augmentation, such as GANs, SMOTE was a better option in this framework to avoid the additional computational requirements associated with the alternatives [31]. After resolving class imbalance,

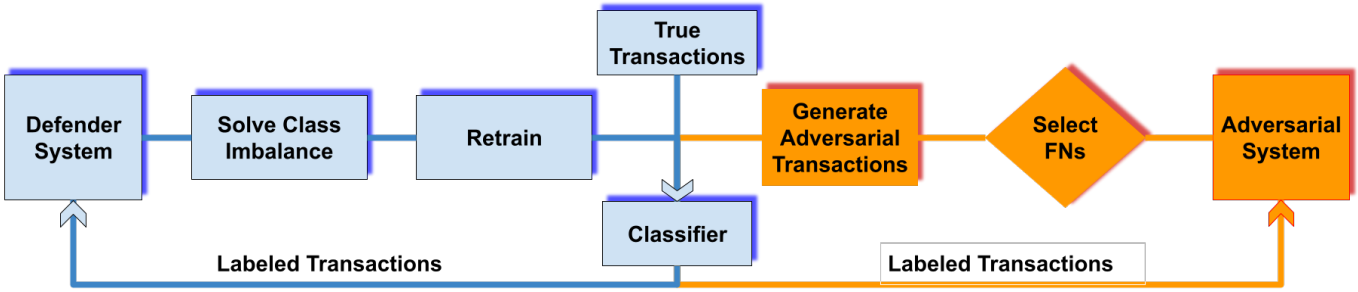


Fig. 2: The overview of the interactions between the adversary and the defender system.

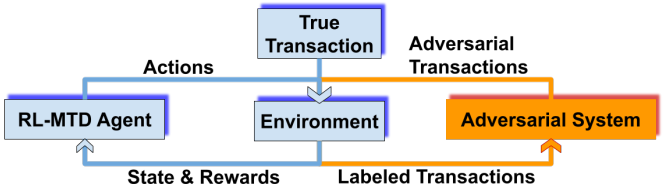


Fig. 3: The procedure of RL-based MTD defense against CCF adversaries

various ML classifiers are trained on each round's labeled data and tested on the next round's new incoming data. We create and test a mixture of defensive strategies in the diverse simulated scenarios, detailed in Section VI-B.

D. RL-based MTD for Credit Card Fraud Detection

Inspired by [11, 35] using RL to identify optimal MTD strategies in the cybersecurity domain, we formulate the RL agent's action, reward, and state as below.

1) **Action Space:** The RL's action space consists of $\mathcal{A} = \{a_0, a_1, a_2, a_3, a_4\}$ where action i, denoted by a_i, refers to classifier i to be used in each round. We set each action to the following classifiers: (1) a_1: Logistic Regression classifier (LR); (2) a_1: Random Forest classifier (RF); (3) a_2: Decision Tree classifier (DT); (4) a_3: AdaBoost classifier; and (5) a_4: XGBoost classifier (XGB). We select these well-known ML classifiers. One may choose other ML classifiers to further enhance the performance of CCF detection.

2) **Reward Function:** The RL agent aims to maximize its accumulate reward, $\mathcal{R}_t$ by taking the best action where $\mathcal{R}_t = \sum_{t=0}^{T} \gamma^t r_t$. Here r_t is an immediate reward at time t, γ is a discount factor, and T is the period of an episode. The agent's immediate reward, r_t, is formulated by:

$$r_t = \alpha * \mathcal{R}^X + \beta * \mathcal{R}^{TD}, \tag{1}$$

where $\mathcal{R}^X$ refers to the reward depending on the outcome of CCF detection (i.e., true positives or true negatives), and $\mathcal{R}^{TD}$ is the reward based on the transfer distance (TD) between the decision spaces of defensive classifiers [36]. Specifically, we define $\mathcal{R}^X$ by:

$$\mathcal{R}^{\mathcal{X}} = \begin{cases} +\frac{X_{11}}{X_{total}} & \text{if } t = 1 \text{ and } p = 1, \\ -\frac{X_{10}}{X_{total}} & \text{if } t = 1 \text{ and } p = 0, \\ +\frac{\lambda X_{00}}{X_{total}} & \text{if } t = 0 \text{ and } p = 0, \\ -\frac{\lambda X_{01}}{X_{total}} & \text{if } t = 0 \text{ and } p = 1, \end{cases} \tag{2}$$

where R^X is ranged as a real number in $[-1, 1]$, p is a predicted transaction label, either 0 or 1, and t represents the true transaction label. X_{total} is the total number of transactions in each round's dataset, while X_{tp} is defined as the segment of data that stands for each (t, p) pair. For instance, X_{11} represents the number of transactions that have a true label of fraud (1), and are correctly predicted by the ML classifier as fraud (1). We also introduce λ as a tunable hyper-parameter between $[0, 1]$, indicating how much the agent should focus on the non-fraud classifications (i.e., true negatives and false positives). The transfer distance (TD), $\mathcal{R}^{TD}$, is given by:

$$\mathcal{R}^{TD} = T_{LC,NC}, \tag{3}$$

where R^{TD} is ranged as a real number in $[0, 1]$, representing the TD between the classifier used in the last round of the iterated game and the new classifier selected by the RL agent. For instance, $T_{LC,NC}$ stands for the TD between the classifier used in the last round, named the *last classifier* (LC) and the newly selected classifier, named the *new classifier* (NC). For example, $T_{1,1}$ refers to $(LC, NC) = (1, 1)$, which means LC is RF and NC is RF, resulting in $T_{1,1} = 0$. The TD value is calculated based on the hamming distance between the predicted fraud/non-fraud arrays of each classifier, and the classifier used last round [37].

3) **State Space:** The state space, $\mathcal{S}$, is defined based on (1) each round's Confusion Matrix data, $\mathbf{C}_M$ after applying λ hyper-parameter in Eq. (2); (2) each round's TD values: $\mathbf{T}_D = (T_{LC,0}, T_{LC,1}, T_{LC,2}, T_{LC,3}, T_{LC,4})$; (3) The current classification model, $\mathcal{M}$; and (4) Number of rounds the current classifier has been consecutively deployed, denoted by $\mathcal{N}_C$. Therefore, we denote $\mathcal{S} = \{\mathbf{C}_M, \mathbf{T}_D, \mathcal{M}, \mathcal{N}_C\}$.

E. Training of RL-MTD Agent

The RL-MTD agent is trained based on a Deep Q-learning Network (DQN), which uses a value-based RL agent to train a

TABLE I: KEY DESIGN PARAMETERS, THEIR MEANINGS, AND DEFAULT VALUES

Notation	Meaning	Default
LDW	Number of rounds between undetected frauds and receiving their delayed label	2
F_R	Percentage of adversarial fraud transactions in each round's data	3 %
ε	Number of RL training epochs	200
R	Number of test rounds in each game	50
λ	RL agent's $\mathcal{R}^{\mathcal{A}}$ reward weight for non-fraud classifications - $[0, 1]$	0.5
α	Reward multiplier for $\mathcal{R}^{\mathcal{A}}$ - real number	6
β	Reward multiplier for $\mathcal{R}^{\mathcal{TD}}$ - real number	8
γ	RL agent's learning discount factor - [0,1]	0.99

Deep Neural Network (DNN) and estimate its rewards for each round using its state-action values. We choose DQN, which is known as the most suitable option to learn the optimal MTD policy in this effort [15, 38]. Our RL agent is trained in an iterative framework, with each training episode composed of a single round of data. Given the DNN nature of the agent, a larger number of datasets is needed for the agent to be trained. Accordingly, we created 100 smaller datasets based on random sampling. Then, the RL agent is trained for 200 episodes, using the 100 smaller datasets twice. Since our agent's reward is normalized based on the total number of transactions in each round, the ranges of the RL-MTD agent's reward and state space values are not affected by the size of the datasets.

During the training phase of the RL-MTD agent, we tested the effect of training two different levels of attack involvement on defense performance where the agent is (1) trained and tested on the same adversary and (2) trained on one adversary and tested on another adversary. Under both cases, our proposed RESONANT does not require prior knowledge of adversaries to produce high effectiveness of its defense because our model uses the model-free design with RL.

VI. EXPERIMENTAL SETUP

This section describes the datasets, comparing defense schemes and metrics used for the conducted experiments.

A. Datasets

The utilized data [31] was provided by a financial institution engaged in the retail banking industry. The dataset was comprised of $\approx$ 80 million anonymized credit card transactions spanning eight months. The dataset contained 70 features, and the transactions were already pre-labeled as fraud or non-fraud. However, to mitigate the curse of high dimensionality in data, we reduced the data dimensionality to 11 features per transaction: (1) Fraud Indicator; (2) Approved Authorization Count; (3) Average Daily Authorization Amount; (4) Merchant Category ID; (5) Point of Service Entry Method ID; (6) Recurring Authorization indicator; (7) Distance From Home; (8) Current Account Balance; (9) Authorization Amount; (10) Authorization Outstanding Amount; and (11) Plastic Issuance Duration.

The data is split into separate subsets to implement the iterated adversarial game framework. A separate dataset is used for each round. The first game is designed to model the interaction between the ML defense and the adversary (see Fig. 2). The second game is tailored to model the interaction between the RL-MTD defense and the adversary (see Fig. 3).

All the data subsets used in the various experiments contain anonymized real-world fraudulent transactions. These transactions serve as real-world seeds for our adversarial fraud generation methods. The dataset can provide realistic fraud transactions in all rounds. Specifically, by the final round, we do not oversample the artificial fraud data created by the previous rounds' oversampling steps.

B. Comparison Defense Mechanisms

We will compare the performance of the following defenses:

- **Static ML**: This uses a single ML classifier with Random Forest (RF), proven the best-performing classifier among the ones considered. We train the RF classifier on the first round's balanced data and then use it to detect fraud across the iterated games without retraining.
- **Naïve ML**: This uses a single ML classifier (i.e., RF) with retraining at the start of each round on the last round's data to keep updated with the changing offensive fraud data trends. Hence, it is an advanced version of the Static ML but incurs extra costs for retraining.
- **Random MTD**: This selects a random classifier each round among the five ML classifiers (i.e., RF, DT, LR, XGB, and AdaBoost), and trains it on the last round's data.
- **RESONANT**: This is the proposed RL-MTD approach using DQN for the model training based on 100 smaller datasets where this scheme is detailed in Section V.

We use the same set of hyperparameters in all defense mechanisms for fair comparison.

C. Metrics

We use the following metrics for our experiments:

- **AUC (Area under the ROC Curve)**: The AUC score is a great way to measure a classification task's performance with respect to true positive and false positive rates. In addition, it is not biased by class imbalances, which is imperative in the CCF detection domain.
- **Attack success ratio (ASR)**: This captures the number of successful fraud transactions over the total number of fraud attempts.
- **Asymptotic complexity in Big-O**: This captures the algorithmic running time in Big-O.

VII. NUMERICAL RESULTS & ANALYSIS

This section demonstrates our experimental results and discusses the underlying reasons for the overall trends in terms of algorithmic complexity analyses, comparative performance analyses, and sensitivity analyses under varying ratios of CCF transactions. For our experimental results in Sections VI-B and VII-C, we use the default values for the key design parameters as described in Table I.

TABLE II: BIG-O COMPLEXITY ANALYSES OF THE FOUR COMPARING SCHEMES

Defense Method	Static ML	Naïve ML	Random MTD	RESONANT
Big-O	$\mathcal{O}\Big(K \times n^2 \times \log(n)\Big)$	$\mathcal{O}\Big(R \times K \times n^2 \times \log(n)\Big)$	$\mathcal{O}\Big(R \times K \times n^2 \times \log(n)\Big)$	$\mathcal{O}\Big(\mathcal{E} \times O(\sum_{l=1}^{2} F_{l-1} \times N_l^2 \times F_l \times \mu_l^2)\Big)$

(Notations: R is the number of rounds in the games. K is the number of variables. n is the number of samples randomly drawn for each tree. ε is the number of DQN training epochs. l is the index of the deep-NN layer. F_{l-1} stands for the number of input channels of the l-th layer. F_l is the number of filters in the l-th layer. N is the size of the filter. μ_l is the size of the output feature map of the l-th layer.)

A. Algorithmic Complexity Analyses

1) **Static ML:** This defense builds an RF model once and then repeatedly deploys it across the iterated rounds of the game without retraining. Hence, we can define its worst-case time complexity as the Big-O complexity of RF. Based on [39], it is given by:

$$\mathcal{O}(\mathcal{K} \times n^2 \times \log(n)), \tag{4}$$

where n is the number of samples, and K is the number of variables randomly drawn at each node.

2) **Naïve ML & Random MTD:** Naïve ML and Random MTD defenses have the following complexity:

$$\mathcal{O}((\mathcal{R} \times K \times n^2 \times \log(n)), \tag{5}$$

where R is the number of rounds in the iterated game, as both defenses fit the ML model in each round. In addition, both defenses have the same complexity because RF has a time complexity that is no less than that of the other models in the MTD framework. Therefore, the MTD will deploy the RF every round in the worst-case scenario.

3) **RESONANT:** The most time consuming step in RESONANT is training the DQN agent. Based on [40, 41], its worst-case time complexity given a fixed sample size n is given by[1]:

$$\mathcal{O}\Big(\mathcal{E} \times \Big(\sum_{l=1}^{2} F_{l-1} \times N_l^2 \times F_l \times \mu_l^2\Big)\Big), \tag{6}$$

where ε is the number of DQN training epochs, l is the index of the deep-NN layer, F_{l-1} stands for the number of input channels of the l-th layer, F_l is the number of filters in the l-th layer, N is the size of the filter, and μ_l is the size of the output feature map of the l-th layer. After the agent is trained, each step of the game could involve re-training with the complexity given in Eq. (5). Therefore, the longer the agent is used, the lower its relative cost.

From our experiments, the DQN agent, RESONANT, takes significantly longer to train than the alternative defense approaches. Yet, the agent does not require retraining between the rounds of the iterative games, so it has a high initial computation cost but then has a complexity matching Eq. (5). Table II summarizes the Big-O complexities of all four schemes compared in this work.

B. Comparative Performance Analyses

Fig. 4 compares the performance of all the considered defenses (i.e., classifiers with or without MTD) in terms of the AUC and ASR curves. To obtain clearer patterns of the results, we used a moving average with a window size of 15. The dotted lines are the actual scores, while the solid lines represent the moving averages. Fig. 4a shows a similar ASR moving average for the Naïve ML, Random MTD, and RESONANT defenses. Yet, it is worth to note the dotted line of the Random MTD shows high fluctuations, representing high unpredictability (i.e., uncertainty) in performance, compared to the RESONANT defense. On the other hand, Fig. 4b shows a clear outperformance of the RESONANT defense in AUC.

We also observe the average and variance of the AUC-ROC and ASR scores in the games in Fig. 4 in Table III. Based on this analysis, we can observe that RESONANT achieves the highest average AUC score while maintaining a relatively low variance in AUC. Specifically, RESONANT results in an AUC variance four times less than that of the Random MTD. On the other hand, Random MTD achieves the lowest ASR average score yet also results in the highest ASR variance. Accordingly, by examining both the average and variance in ASR, RESONANT produces the second lowest ASR average while having a three times less variance than that of Random MTD.

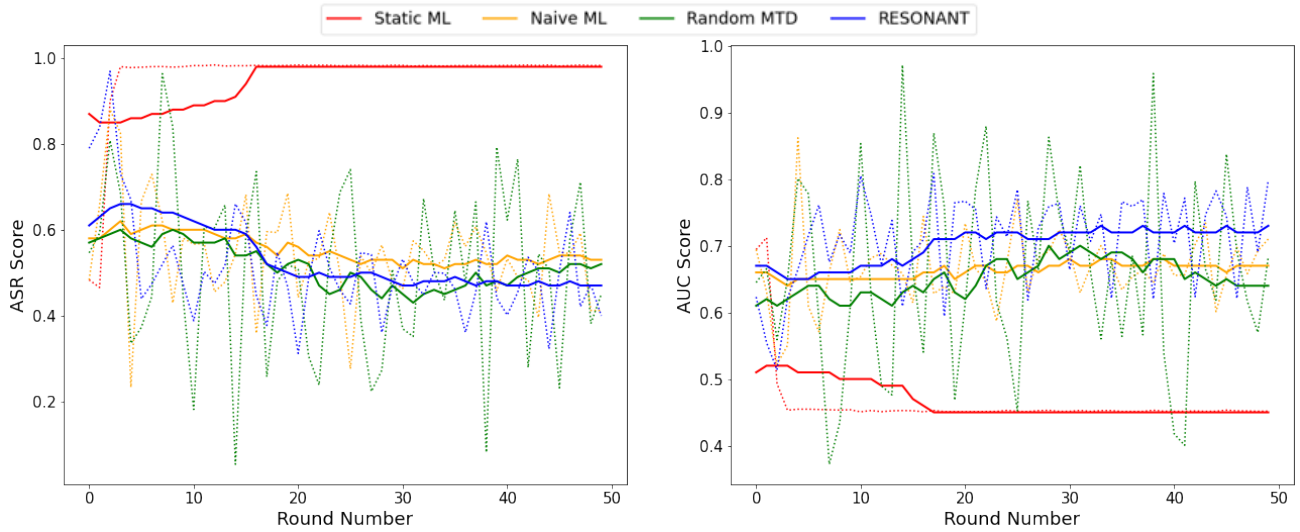
C. Effect of Varying the CCF Rate

Fig. 5 analyzes the effect of varying the degree of attempted CCF transactions in percentage on ASR and AUC. Higher CCF rate represents higher severity in CCF attacks. Fig. 5a shows that RESONANT performs comparably as Naïve ML and Random MTD when the CCF rate is relatively low (i.e., $< 4\%$) while it outperforms under higher CCF rates (i.e., ≥ 4) in ASR. Fig. 5b demonstrates the clear out-performance of RESONANT overall (i.e., $\geq$ 2 %). A moving average with a window size of 30 was used to clearly obtain the overall patterns of the results.

D. Effect of Varying the Size of Label Delay Windows

In Fig. 6, we also investigate the performance of different defenses against CCF attacks while varying the size of the label delay window. The label delay window refers to the number of rounds between an occurrence of undetected fraud and receiving its truth label, representing the window between fraud occurrences and receiving complaints from the CC owners. A larger label delay window represents a harsher condition to the defender because having more delayed access to the truth labels makes it harder for the defender to make accurate fraud detection decisions. Moreover, we utilized a moving average with a window size of 50 to clearly obtain the overall patterns of the results. Fig. 6a shows a comparable performance of the Random MTD and RESONANT defenses

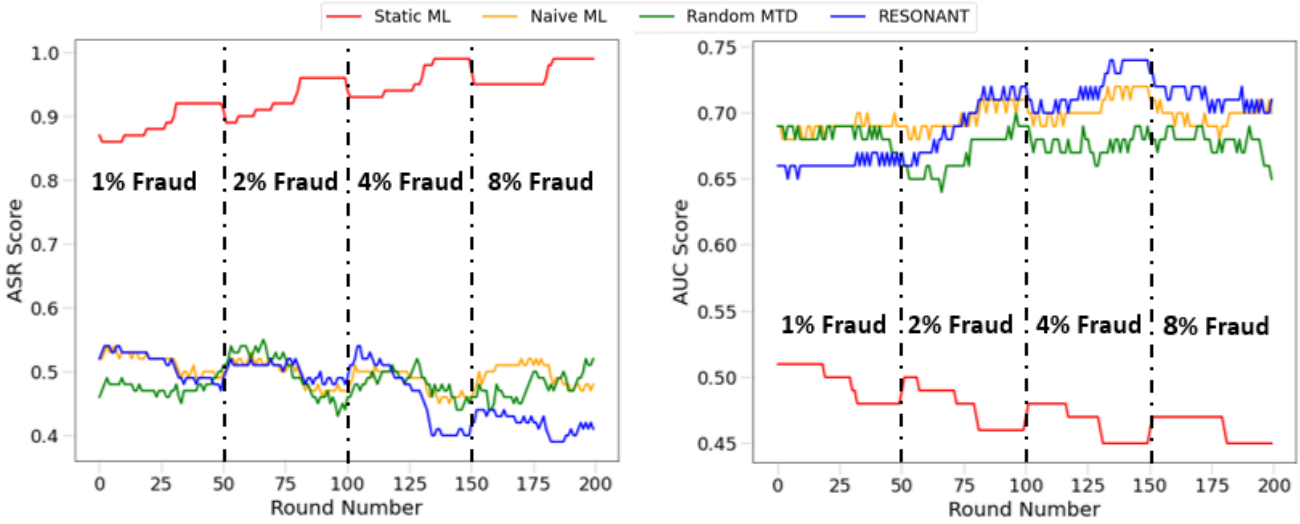
[1]Here given for DQN agent's that use convolutional neural networks (CNN) even though we use fully connected layers.



(a) ASR Under 3% CCF Transactions and 2 Round LDW

(b) AUC Under 3% CCF Transactions and 2 Round LDW

Fig. 4: Comparative Performance Analyses of Static ML, Naïve ML, Random MTD, and RESONANT in ASR and AUC Under 3% CCF Transactions and 2 Round Label Delay Window (LDW).



(a) ASR Under Varying the Percentage of CCF Transactions

(b) ASR Under Varying the Percentage of CCF Transactions

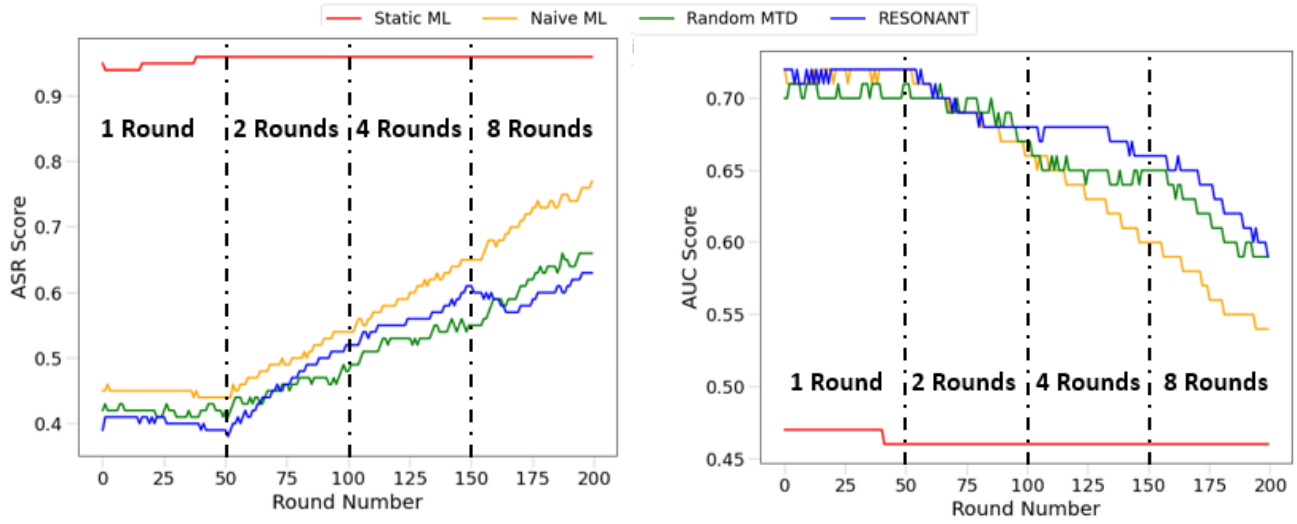
Fig. 5: Comparative Performance Analyses of Static ML, Naïve ML, Random MTD, and RESONANT in ASR and AUC Under Varying the Percentage of CCF Transactions.

in ASR. RESONANT outperforms the other defenses at the 1-round and 8-round delay windows, while the Random MTD outperforms at the 2-round and 4-round delay windows. On the other hand, Fig. 6b illustrates a tied performance between the Naïve ML and RESONANT defenses at the lower fraud rates. Yet, there is a clear outperformance of the RESONANT defense under the harsher adversarial conditions, starting at a 4-round label delay window. Test sets utilized in the games include more than 10 million transactions. Thus, improvements in AUC scores can represent significantly large monetary gains to the defender.

VIII. Conclusions & Future Work

While prior defense mechanisms against adversarial examples aim to optimize a fixed target defense, this effort aims to develop a novel model-free MTD framework based on the autonomous nature of RL in identifying optimal solution(s). We recapitulate the **key contributions** of the proposed credit card fraud (CCF) detection as follows:

- We proposed a novel MTD strategy to bypass vulnerabilities and risks introduced by using the static nature of most current defenses.
- We designed a deep-RL agent named RESONANT to create



(a) ASR Under Varying LDW

(b) ASR Under Varying LDW

Fig. 6: Comparative Performance Analyses of Static ML, Naïve ML, Random MTD, and RESONANT in ASR and AUC Under Varying Label Delay Windows (LDW).

TABLE III: PERFORMANCE SUMMARY OF THE ITERATED ADVERSARIAL GAMES

Defense Strategy	Average AUC	Variance in AUC	Average ASR	Variance in ASR
Static ML	0.463	**0.002**	0.96	**0.010**
Naïve ML	0.666	0.003	0.543	0.014
Random MTD	0.656	0.02	**0.504**	0.041
RESONANT	**0.708**	0.005	0.512	0.016

a model-free defense for an RL agent to autonomously identify the MTD's optimal triggering conditions.
- We modeled attack-defense interactions in a game-theoretic sense to reflect realistic scenarios. This was realized by employing the iterative interaction between the fraudsters and the defender (finance companies) in the real world, representing more realistic representations of real-world adversarial capabilities in the CCF contexts.

Our experimental study obtained the following **key findings**:

- The proposed defense method results in a more effective and consistent defensive performance over time, exhibiting a higher average performance with a lower variance.
- Using `RESONANT` is more beneficial in harsher conditions than friendly conditions, such as higher rates of CCF or longer delays to obtain the truth information (i.e., longer label delay windows).

For **future research**, we plan to (1) experiment with other deep-RL agents in search of the optimal MTD decision making agent; (2) expand the action space of the agent to include manipulating both which ML model to use and what hyper-parameters to use for a particular model; and (3) utilize a DNN-based artificial fraud generation method to explore the performance of `RESONANT` against a DNN-based adversarial model and compare it to that of other defense techniques.

REFERENCES

[1] S. Vimal, K. Kayathwal, H. Wadhwa, and G. Dhama, "Application of deep reinforcement learning to payment fraud," Dec. 2021.

[2] A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," in *2007 International Conference on Service Systems and Service Management*, Jun. 2007, pp. 1–4.

[3] M. F. Zeager, A. Sridhar, N. Fogal, S. Adams, D. E. Brown, and P. A. Beling, "Adversarial learning in credit card fraud detection," in *2017 Systems and Information Engineering Design Symposium (SIEDS)*, Apr. 2017, pp. 112–116.

[4] R. Colbaugh and K. Glass, "Predictive moving target defense." Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), Tech. Rep., May 2012.

[5] C. Lei, H.-Q. Zhang, J.-L. Tan, Y.-C. Zhang, and X.-H. Liu, "Moving target defense techniques: A survey," *Security and Communication Networks*, vol. 2018, Jul. 2018.

[6] I. Goodfellow, "A research agenda: Dynamic models to defend against correlated attacks," *arXiv preprint arXiv:1903.06293*, Mar. 2019.

[7] Y. Hong, U. Hwang, J. Yoo, and S. Yoon, "How generative adversarial networks and their variants work: An

overview," *ACM Computing Surveys (CSUR)*, vol. 52, no. 1, pp. 1–43, Feb. 2019.

[8] S. Arora, R. Ge, Y. Liang, T. Ma, and Y. Zhang, "Generalization and equilibrium in generative adversarial nets (gans)," in *International Conference on Machine Learning*, Jul. 2017, pp. 224–232.

[9] J.-H. Cho, D. P. Sharma, H. Alavizadeh, S. Yoon, N. Ben-Asher, T. J. Moore, D. S. Kim, H. Lim, and F. F. Nelson, "Toward proactive, adaptive defense: A survey on moving target defense," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 709–745, Jan. 2020.

[10] A. Chowdhary, D. Huang, A. Sabur, N. Vadnere, M. Kang, and B. Montrose, "Sdn-based moving target defense using multi-agent reinforcement learning," in *Proceedings of the first International Conference on Autonomous Intelligent Cyber defense Agents (AICA 2021), Paris, France*, Mar. 2021, pp. 15–16.

[11] T. Eghtesad, Y. Vorobeychik, and A. Laszka, "Adversarial deep reinforcement learning based adaptive moving target defense," in *Decision and Game Theory for Security: 11th International Conference, GameSec 2020, College Park, MD, USA, October 28–30, 2020, Proceedings 11*, Oct. 2020, pp. 58–79.

[12] ——, "Deep reinforcement learning based adaptive moving target defense," *arXiv preprint arXiv:1911.11972*, Nov. 2019.

[13] M. Zhu, Z. Hu, and P. Liu, "Reinforcement learning algorithms for adaptive cyber defense against heartbleed," in *Proceedings of the first ACM Workshop on Moving Target Defense*, Nov. 2014, pp. 51–58.

[14] S. Yoon, J.-H. Cho, D. S. Kim, T. J. Moore, F. Free-Nelson, and H. Lim, "DESOLATER: Deep reinforcement learning-based resource allocation and moving target defense deployment framework," *IEEE Access*, vol. 9, pp. 70 700–70 714, Apr. 2021.

[15] X. Chai, Y. Wang, C. Yan, Y. Zhao, W. Chen, and X. Wang, "DQ-MOTAG: deep reinforcement learning-based moving target defense against ddos attacks," in *2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC)*. IEEE, Jul. 2020, pp. 375–379.

[16] S. Kim, S. Yoon, J.-H. Cho, D. S. Kim, T. J. Moore, F. Free-Nelson, and H. Lim, "Divergence: deep reinforcement learning-based adaptive traffic inspection and moving target defense countermeasure framework," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4834–4846, Jan. 2022.

[17] L. Zhinin-Vera, O. Chang, R. Valencia-Ramos, R. Velastegui, G. E. Pilliza, and F. Quinga-Socasi, "Q-credit card fraud detector for imbalanced classification using reinforcement learning." in *ICAART (1)*, 2020, pp. 279–286.

[18] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 631–641, Jan. 2019.

[19] S. Khatri, A. Arora, and A. P. Agrawal, "Supervised machine learning algorithms for credit card fraud detection: a comparison," in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2020, pp. 680–683.

[20] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *2017 International Conference on Computing Networking and Informatics (ICCNI)*, Oct. 2017, pp. 1–9.

[21] S. Dhankhad, E. Mohammed, and B. Far, "Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, Jul. 2018, pp. 122–125.

[22] K. Chaudhary, J. Yadav, and B. Mallick, "A review of fraud detection techniques: Credit card," *International Journal of Computer Applications*, vol. 45, no. 1, pp. 39–44, May. 2012.

[23] X. Zhang, Y. Han, W. Xu, and Q. Wang, "Hoba: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture," *Information Sciences*, vol. 557, pp. 302–316, May. 2021.

[24] A. Chakraborty, M. Alam, V. Dey, A. Chattopadhyay, and D. Mukhopadhyay, "A survey on adversarial attacks and defences," *CAAI Transactions on Intelligence Technology*, vol. 6, no. 1, pp. 25–45, Mar. 2021.

[25] ——, "Adversarial attacks and defences: A survey," *arXiv preprint arXiv:1810.00069*, Sep. 2018.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[27] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[28] S. Sankaranarayanan, A. Jain, R. Chellappa, and S. N. Lim, "Regularizing deep networks using efficient layerwise adversarial training," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[29] T. Cody, S. Adams, and P. A. Beling, "A utilitarian approach to adversarial learning in credit card fraud detection," in *2018 Systems and Information Engineering Design Symposium (SIEDS)*, Apr. 2018, pp. 237–242.

[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002.

[31] A. Langevin, T. Cody, S. Adams, and P. Beling, "Generative adversarial networks for data augmentation and transfer in credit card fraud detection," *Journal of the Operational Research Society*, vol. 73, no. 1, pp. 153–180, Jan. 2022.

[32] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th International Conference on Information and*

Communication Systems (ICICS), Apr. 2020, pp. 243–248.

[33] E. Lin, Q. Chen, and X. Qi, "Deep reinforcement learning for imbalanced classification," *Applied Intelligence*, vol. 50, pp. 2488–2502, Aug. 2020.

[34] J. Fan, Z. Wang, Y. Xie, and Z. Yang, "A theoretical analysis of deep q-learning," in *Learning for Dynamics and Control*, Jul. 2020, pp. 486–489.

[35] S. Sengupta and S. Kambhampati, "Multi-agent reinforcement learning in bayesian stackelberg markov games for adaptive moving target defense," *arXiv preprint arXiv:2007.10457*, Jul. 2020.

[36] T. Cody, S. Adams, and P. A. Beling, "Empirically measuring transfer distance for system design and operation," *IEEE Systems Journal*, vol. 16, no. 3, pp. 4962–4973, Feb. 2022.

[37] M. Norouzi, D. J. Fleet, and R. R. Salakhutdinov, "Hamming distance metric learning," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[38] A. Jeerige, D. Bein, and A. Verma, "Comparison of deep reinforcement learning approaches for intelligent game playing," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2019, pp. 0366–0371.

[39] G. Louppe, "Understanding random forests: From theory to practice," *arXiv preprint arXiv:1407.7502*, Jul. 2014.

[40] N. Gao, Z. Qin, X. Jing, Q. Ni, and S. Jin, "Anti-intelligent uav jamming strategy via deep q-networks," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 569–581, Oct. 2019.

[41] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5353–5360.