Hindawi
Security and Communication Networks
Volume 2020, Article ID 8824163, 10 pages
https://doi.org/10.1155/2020/8824163

Research Article

Game Theoretical Method for Anomaly-Based Intrusion Detection

Zhiyong Wang,[1] Shengwei Xu,[2] Guoai Xu,[1] Yongfeng Yin,[3] Miao Zhang,[1] and Dawei Sun[4]

[1]*National Engineering Laboratory of Mobile Network Security, School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, China*
[2]*Deputy Director of Information Security Research Institute, Beijing Institute of Electronic Science and Technology, Beijing, China*
[3]*School of Reliability and Systems Engineering, Beihang University, Beijing, China*
[4]*R&D Department, Beijing Softsec Technologies Co., Ltd., Beijing, China*

Correspondence should be addressed to Yongfeng Yin; yyf@buaa.edu.cn

Received 8 May 2020; Revised 24 June 2020; Accepted 19 August 2020; Published 4 September 2020

Academic Editor: Xiaolong Xu

In this paper, the game theoretical analysis method is presented to provide optimal strategies for anomaly-based intrusion detection systems (A-IDS). A two-stage game model is established to represent the interactions between the attackers and defenders. In the first stage, the players decide to do actions or keep silence, and in the second stage, attack intensity and detection threshold are considered as two important strategic variables for the attackers and defenders, respectively. The existence, uniqueness, and explicit computation of the Nash equilibrium are analyzed and obtained by considering six different scenarios, from which the optimal detection and attack actions are provided. Numerical examples are provided to validate our theoretical results.

1. Introduction

Nowadays, network devices and communication services are vulnerable to various kinds of intrusion attacks, such as DoS/DDoS, false data injection, and botnet attacks. The intrusion attacks tend to be more intelligent and the unexpected attack modes arise frequently. Consequently, great challenges are brought into network security control and management. As one of the most important techniques to tackle with various attacks, anomaly-based intrusion detection system (A-IDS) has been widely adopted in almost all kinds of network environments [1, 2]. An anomaly-based intrusion detector attempts to estimate the normal behavior with a profile and generates an anomaly alarm once the profile collected from real-time observation exceeds a predefined threshold [2].

In an intrusion detection system, the attacker and defender can naturally be regarded as two players who try to maximize their payoffs, respectively, by executing certain optimal strategies. Thus, the game theoretical method is an effective tool which enables a defender to earn the maximum payoff (or the minimum loss) while fighting with the attacks. A number of results on game theory-based intrusion detection methods have been reported for different network environments and security requirements. Excellent surveys about this topic can be found in [3–6]. In [7], two-player noncooperative strategic game models are established for some general intrusion detection problems and Nash equilibriums are analyzed explicitly. In [8–11], game theoretical intrusion detection methods are investigated to solve the security resource allocation problems of large-scale heterogeneous networks. Note that, in [8–10], it is assumed that the defender scan always correctly identify the malicious behaviors of the attackers without any errors, while such an assumption may not be satisfied in some cases. For example, for intelligent APT attacks, the attackers often disguise themselves as no attack happens, which may make the detector to not always preciously identify the malicious actions. To handle these uncertainties, Bayesian games are

considered in intrusion detection by updating the defender's belief to her/his opponent based on the past behaviors [12–15]. The main idea of Bayesian game-based intrusion detection is to use probability to represent the uncertainties and further use Bayesian iteration to update the dynamics. For self-organizing ad hoc networks, some nodes may be malicious and how to detect the malicious actions is an important work. Some strategic games are presented to stimulate the cooperation among distinct regular nodes, based on which the hidden malicious nodes can be detected [16–21]. In [22], a two-player Stackelberg stochastic game is analyzed for achieving the best response against the intrusion. In [23, 24], game theory-based analysis methods for distributed intrusion detection are proposed, where consensus-based distributed detection method is presented and then game analysis is provided for the optimal defense and attack strategies. In [25], the privacy defense problem is also considered in the collaborative security scheme design problem by using the game theoretical analysis method. In [26], a differential game model is established to analyze the dynamic process of the attack and defense.

In a game between an attacker and a defender, the rational attacker will not launch an attack otherwise she/he can get a positive payoff. Moreover, the attack intensity needs be chosen to maximize her/his positive payoff. On the contrary, the defender will perform a defense action to resist the attack according to a similar rule. In an A-IDS, a predefined detection threshold needs be cautiously determined. In general, a higher threshold with a larger normal coverage area will result in a smaller false alarm rate but a larger missing alarm rate. Note that the missing alarm rate is also closely related to the attack intensity. More specifically, larger attack intensity will cause a lower missing alarm rate. Though attack intensity and detection threshold are two important factors affecting the false and missing alarm rates, which correspond to the payoffs of attackers and defenders in an intrusion detection game, they are seldom considered in the aforementioned results. In most of the aforementioned works, the false and missing alarm rates are assumed to be known constants and only binary actions "do" or "not do" are considered in their game models. In [11], the detection threshold and attack intensity are considered, while the focus is mainly related to distributed resource allocation of the heterogeneous networks.

Motivated by the limitations mentioned above in the literature, a more realistic two-stage form of the intrusion detection game model is presented in this paper. The attack intensity and detection threshold are considered as two strategic variables. In the first stage, the attackers and defenders make decisions on whether the attack and defense actions should be executed, respectively. Once the attack/monitoring actions are decided to be executed, optimal attack intensity and detection threshold are determined to maximize their utilities in the second stage. The existence and uniqueness of the Nash equilibrium are discussed for the first stage of our presented game model under different scenarios, when the strategic variables of the second stage are restricted to certain regions. Then, the optimal attack intensity and detection threshold are derived for each scenario, correspondingly.

The contributions of this paper can be summarized as follows:

(1) A two-stage game model is presented for anomaly-based intrusion detection confrontation. In contrast to the existing work, where only binary actions "do" or "not do" are considered in the game model, the attack intensity and the detection threshold are considered as two key strategic variables, and the false and missing alarm rates are the functions of the attack intensity and the detection threshold, instead of being assumed to be constant. The two stages of the game model are tightly coupled with each other and thus the game model is more complex.

(2) The existence, uniqueness, and calculation of Nash equilibriums are discussed. Based on the results, optimal selections of attack intensity and detection threshold for achieving the maximum payoffs of the attackers and defenders are provided. The results provide a new method to determine the detection threshold in the defense, from the perspectives of the optimization and confrontation. So, the presented game model and Nash equilibrium solution give a more realistic theoretical analysis framework for the anomaly-based security detection.

The rest of this paper is organized as follows. In Section 2, some definitions are introduced and a two-stage game model of the A-IDS is presented. In Section 3, the Nash equilibrium of the proposed game model is analyzed. Simulation results are given to show the effectiveness of our game theoretical analysis methods in Section 4, followed by the conclusions of the paper summarized in Section 5.

2. A Two-Stage Intrusion Detection Game Model

Suppose that there is a network unit vulnerable to intrusion attacks. Typical examples for such a unit include a software system, network equipment, and a communication channel. Here, we adopt similar attack and A-IDS detection models as that in [11]. The strategic form of two-player noncooperative game is given in Table 1. U_A and U_D denote the payoffs of the attacker and the defender, respectively.

In the following, we give the physical meanings of the corresponding variables in Table 1. The variable x denotes the attack intensity, for example, the number of attack packets in a DoS/DDoS attack, or the number of bogus packets in a DNS cache poisoning attack or jamming strength in a communication attack, or the magnitude of false data injection. It is assumed that $x \in [\underline{x}, \overline{x}]$, where $1 \geq \overline{x} > \underline{x} > 0$. The function $s(x) \in \mathfrak{R}$ is used to represent the extent of damage to the security of the unit, when it is suffered from an attack with intensity x. It is natural to consider $s(x)$ as a strictly increasing function such that $(\partial s(x)/\partial x) > 0$ and $s(x) \in [\underline{s}, \overline{s}]$ with $1 \geq \overline{s} > \underline{s} > 0$. The term $c_1 W + u(x)W$, where $c_1 \in (0, 1)$ is a constant, W is the security asset of the unit, and $u(x)$ is a strictly increasing function, denotes the cost of launching the attack. The

Table 1: Strategic form of the local game.

	Monitor	Not monitor
Attack	$U_A(x,y) = q(x,y)s(x)W - c_1W - u(x)W$, $U_D(x,y) = -q(x,y)s(x)W - c_2W$	$U_A(x,y) = s(x)W - c_1W - u(x)W$, $U_D(x,y) = -s(x)W$
Not attack	$U_A(x,y) = 0$, $U_D(x,y) = -p(y)c_3W - c_2W$	$U_A(x,y) = 0$, $U_D(x,y) = 0$

variable y denotes the detection threshold. It is assumed that $y \in (\underline{y}, \overline{y})$ with $\overline{y} > \underline{y} > 0$ and a larger y corresponds to a larger coverage area for normal behavior. The function p denotes the false alarm rate, i.e., it represents the probability that an alarm is generated though no attack is activated. Obviously, p is determined completely by the threshold y and $p(y)$ is a strictly decreasing function in this paper. The function q denotes the missing alarm rate, i.e., it represents the probability that no alarm is generated though an attack is executed. The function q is determined by both attack intensity x and threshold y. It can be easily derived that q is strictly decreasing and increasing with respect to x and y, respectively. The parameters $c_2 \in (0,1)$ and $c_3 \in (0,1)$ are two constants.

Clearly, the game model described in Table 1 contains the following two stages. In the first stage, the optimal strategy set "Attack/Not attack" and "Monitor/Not monitor" needs be determined by the attacker and defender. Then, both players proceed to the next stage to select optimal attack intensity x and detection threshold y. For better understanding, the two-stage pure-strategic intrusion detection game model with one attacker and one defender is described in Table 2 in a more rigorous way.

Remark 1. The attack and detection models are similar to that in [11], while the results of [11] mainly consider the attack and defense resource allocation problem for heterogeneous distributed networks. In this paper, we consider the confrontation problem for one network unit, as expressed by the game model in Table 2. Thus, it is essentially different from the work in [11]. Besides, we establish a two-stage game model by considering the attack intensity and detection threshold as the key strategic variables, which is also different from the existing works.

3. Nash Equilibrium Analysis of the Game

As mentioned in Section 2, the attacker/defender needs to decide whether to launch an attack/to monitor the unit or keep silence in the first stage of the presented game model. For simplicity, an extra assumption is imposed that if the payoffs of a player choosing to perform the action and to keep silence are the same, she/he will keep silence. In other words, the attacker/defender tends to do nothing if she/he cannot earn larger payoffs by launching an attack/monitoring. Note that the value of W has no impact on the analysis of Nash equilibrium (*hereinafter referred to as NE*) of the game from Table 1. Thus, without loss of generality, we set $W = 1$.

Denote the feasible set of x and y by π with $\pi \in [\underline{x}, \overline{x}] \times [\underline{y}, \overline{y}]$. For convenience in later analysis, π is divided into the following subsets:

$$\begin{aligned}
\pi_1 &= \{(x,y) \in \pi:\ s(x) - c_1 - u(x) \le 0\},\\
\pi_2 &= \{(x,y) \in \pi:\ s(x) - c_1 - u(x) \ge 0, -q(x,y)s(x) - c_2 \le -s(x)\},\\
\pi_3 &= \{(x,y) \in \pi:\ q(x,y)s(x) - c_1 - u(x) \ge 0, -q(x,y)s(x) - c_2 \ge -s(x)\},\\
\pi_4 &= \{(x,y) \in \pi:\ q(x,y)s(x) - c_1 - u(x) < 0, s(x) - c_1 - u(x) > 0, -q(x,y)s(x) - c_2 > -s(x)\}.
\end{aligned} \tag{1}$$

It can be readily shown that $\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 = \pi$ and $(\pi_1 \cup \pi_2 \cup \pi_3) \cap \pi_4 = \varnothing$. The results of NE for the game as described in Tables 1 and 2 will be obtained from the following scenarios. In Scenario L.1, only one subset of π_1, π_2, π_3, and π_4 is nonempty. In Scenarios L.2~L.5, π_4 is empty while at least two subsets of π_1, π_2, and π_3 are nonempty. In Scenario L.6, π_4 and at least one subset of π_1, π_2, and π_3 are nonempty. Clearly, there is no overlap between any two scenarios and the six scenarios include all the possibilities. In the following, the sufficient and necessary conditions on x and y for the existence and uniqueness of NE are first derived for Scenarios L.1~L.6, respectively. Then, the optimal values of x and y, denoted by x^* and y^*, are further provided.

For convenient expression in what follows, two variables x' and x'' are first defined, i.e.,

$$x' = \arg\max_x q\left(x, \underline{y}\right)s(x) - u(x) - c_1, \quad \text{s.t. } x \in \left[\underline{x}, \overline{x}\right], \tag{2}$$

$$x'' = \arg\max_x s(x) - c_1 - u(x), \quad \text{s.t. } x \in [\underline{x}, \overline{x}]. \tag{3}$$

The optimization problems presented by (2) and (3) can be solved by classical optimization methods such as the gradient method and Lagrangian multiplier method [27].

Table 2: Two-stage pure-strategic intrusion detection game.

Players	Attacker, defender
Strategy sets	Attacker: Attack, not attack, attack intensity x Defender: Monitor, not monitor, detection threshold y
Constraints	$x \in [\underline{x}, \overline{x}]$, $\overline{x} > \underline{x} > 0$, $y \in [\underline{y}, \overline{y}]$, $\overline{y} > \underline{y} > 0$, $s(x) \in [\underline{s}, \overline{s}]$, $1 \geq \overline{s} > \underline{s} > 0$, $(\partial s(x)/\partial x) > 0$, $(\partial u(x)/\partial x) > 0$, $(\partial \overline{p}(y)/\partial y) < 0$, $(\partial q(x, y)/\partial x) < 0$, $(\partial q(x, y)/\partial y) > 0$, $p \in [p, \overline{p}]$, $q \in [q, \overline{q}]$, $1 > \overline{p} > p > 0$, $1 > \overline{q} > q > 0$, $c_1, c_2, c_3 \in (0, 1)$
Payoffs	U_A, U_D (see Table 1)
Game target	The players choose their strategies to maximize their payoffs U_A, U_D

Scenario L.1. Only one of the subsets π_1, π_2, π_3, and π_4 is nonempty.

The following conclusions can be drawn.

Theorem 1. *In Scenario L.1, the NE of the game, as described in Table 1, is derived as follows:*

(1) If only the subset $\pi_1 \neq \emptyset$, "not attack, not monitor" is the unique NE

(2) If only the subset $\pi_2 \neq \emptyset$, "attack, not monitor" is the unique NE and $x^* = x''$

(3) If only the subset $\pi_3 \neq \emptyset$, "attack, monitor" is the unique NE and $x^* = x'$, $y^* = \underline{y}$

(4) If only the subset $\pi_4 \neq \emptyset$, no NE exists

Proof. Firstly, the strategy combination "attack, not monitor" will not be the NE. This is because, $-p(y)c_3 - c_2 < 0$, the defender tends to "not monitor" the unit to earn zero payoff:

(1) If only $\pi_1 \neq \emptyset$, we have $q(x, y)s(x) - c_1 - u(x) < s(x) - c_1 - u(x) \leq 0$. This indicates that the attacker has no incentive to launch an attack either. Therefore, "not attack, not monitor" is the unique NE.

(2) If only $\pi_2 \neq \emptyset$, as the payoff of the attacker $s(x) - c_1 - u(x)$ is positive for any attack intensity x, the attacker will select "attack." Besides, the defender will never get more payoffs when she/he selects "monitor" as $-q(x, y)s(x) - c_2 \leq -s(x)$ for an arbitrary threshold y. Thus, the defender will select "monitor." The optimal attack intensity x^* should be derived by maximizing the payoff of the attack; therefore, $x^* = x''$ based on (3).

(3) If only $\pi_3 \neq \emptyset$, the attacker will always select "attack." This is because for any attack intensity x and detection threshold y the payoff of the attacker satisfies $s(x) - c_1 - u(x) > q(x, y)s(x) - c_1 - u(x) \geq 0$. Since the payoff of the defender satisfies $-q(x, y)s(x) - c_2 \geq -s(x)$ for an arbitrary y, the defender will select "monitor." Then, for the defender, the optimal threshold is computed by

$$y^* = \arg\max_y -q(x, y)s(x) - c_2, \quad \text{s.t. } x \in [\underline{x}, \overline{x}], y \in [\underline{y}, \overline{y}]. \tag{4}$$

Based on the property that $(\partial q(x, y)/\partial y) > 0$ in Table 2, we have $y^* = \underline{y}$. Then, the optimal attack intensity is given by $x^* = x'$ based on (2).

(4) If only $\pi_4 \neq \emptyset$, "attack, monitor" cannot be the NE since $q(x, y)s(x) - c_1 - u(x) < 0$. Meanwhile, "attack, not monitor" is not the NE because $s(x) - q(x, y)s(x) > c_2$ indicates that the defender will selects "monitor".Moreover,since $s(x) - c_1 - u(x) > 0$, "not attack, not monitor" cannot be the NE, either. Combining with the result derived in the beginning that "not attack, monitor" cannot be the NE, it is concluded that no NE exists.

Remark 2. From Theorem 1, the payoffs of the two players are, respectively, expressed as $U_A = s(x^*) - c_1 - u(x^*)$ and $U_D = -s(x^*)$ in (2) in Scenario L.1. It implies that the attacker obtains positive payoff while the defender loses certain security asset in this scenario. On the contrary, the payoffs of two players are, respectively, expressed as $U_A = q(x^*, \underline{y})s(x^*) - u(x^*) - c_1$ and $U_D = -q(x^*, \underline{y})s(x^*) - c_2$ in (3) in Scenario L.1. Similar to (2) in Scenario L.1, the attacker earns positive payoff while the defender loses certain security asset. Nevertheless, different from (2) in Scenario L.1, the defender compensates for part of the loss by executing monitoring action in this scenario as $q < 1$. Thus, the payoff earned by the attacker decreases.

As discussed previously, Scenarios L.2~L.5 cover the possibilities that π_4 is empty while at least two subsets of π_1, π_2, and π_3 are nonempty. Details are given as below.

Scenario L.2. $\pi_1 \neq \emptyset$, $\pi_2 \neq \emptyset$, and $\pi_3 = \pi_4 = \emptyset$.

The following results about the NE for this scenario can be shown.

Theorem 2. *In Scenario L.2, the strategy combination "attack, not monitor" is the unique NE and $x^* = x''$.*

Proof. The subset $\pi_2 \neq \emptyset$ indicates that there exists an x such that the payoff of the attacker $s(x) - u(x) - c_1$ is positive. Thus, the attacker will select the strategy "attack." Besides, the payoff of the defender satisfies $-q(x, y)s(x) - c_2 \leq -s(x)$ for any threshold y, so the defender will select "not monitor." Besides, the optimal attack intensity is given by $x^* = x''$. □

Scenario L.3. $\pi_1 \neq \varnothing$, $\pi_3 \neq \varnothing$, and $\pi_2 = \pi_4 = \varnothing$.

Main results for this scenario are formally stated in the following theorem.

Theorem 3. *In Scenario L.3, the strategy combination "attack, monitor" is the unique NE if and only if* $q(x', \underline{y})s(x') - c_1 - u(x') > 0$. *The optimal attack intensity and detection threshold are* $x^* = x'$ *and* $y^* = \underline{y}$.

Proof. Necessity: if "attack, monitor" is the unique NE, then from (2) and (4), there are $x^* = x'$ and $y^* = \underline{y}$. The payoff of the attacker with x^* and y^* must be positive; thus, $q(x', \underline{y})s(x') - c_1 - u(x') > 0$.

Sufficiency: since $q(x', \underline{y})s(x') - c_1 - u(x') > 0$, the attacker can earn a positive maximum payoff if the defender selects the strategy "monitor" and $y^* = \underline{y}$. Thus, the attacker will select to "attack" and $y^* = \underline{y}$. As $q < 1$ and $(\partial q(x, y)/\partial y) > 0$, there is $s(x') - c_1 - u(x') > q(x', y)s(x') - c_1 - u(x') \geq q(x', \underline{y})s(x') - c_1 - u(x') > 0$ for $y \in [\underline{y}, \overline{y}]$. It follows that $x' \notin \pi_1$ and $(x', y) \in \pi_3$ for $y \in [\underline{y}, \overline{y}]$. From the definition of π_3, it can be concluded that $-q(x', y)s(x') - c_2 \geq -s(x')$ for $y \in [\underline{y}, \overline{y}]$. This indicates that no matter how the threshold is selected, the defender will earn larger payoff when she/he selects the strategy "monitor" rather than "not monitor." Clearly, the defender will select "monitor" and the optimal threshold is set as $y^* = \underline{y}$ from (4). Therefore, the strategy combination "attack, monitor" is the unique NE and $x^* = x'$ and $y^* = \underline{y}$. □

Scenario L.4. $\pi_2 \neq \varnothing$, $\pi_3 \neq \varnothing$, and $\pi_1 = \pi_4 = \varnothing$.

The following conclusions can be drawn for this scenario.

Theorem 4. *In Scenario L.4,*

(1) If and only if $-q(x'', \underline{y})s(x'') - c_2 \leq -s(x'')$, *"attack, not monitor" is the NE and* $x^* = x''$

(2) If and only if $-q(x', \underline{y})s(x') - c_2 > -s(x')$, *"attack, monitor" is the NE and* $x^* = x'$ *and* $y^* = \underline{y}$

Proof

(1) Necessity: under the strategy combination "attack, not monitor", the attacker will select x'' as the optimal attack intensity. If $-q(x'', \underline{y})s(x'') - c_2 > -s(x'')$, the defender will select "monitor" to earn larger payoffs, which is a contradiction to the premise that "attack, not monitor" is the NE. Thus, the necessity is shown.

Sufficiency: from the definitions of π_2 and π_3, the attacker can always earn positive maximum payoff when s/he selects "attack." As $\partial q/\partial y > 0$, there is

$$\begin{aligned} &-q(x'', y)s(x'') - c_2 \\ &\quad \leq -q(x'', \underline{y})s(x'') - c_2 \\ &\quad \leq -s(x''). \end{aligned} \tag{5}$$

This means when the attacker selects $x^* = x''$, the defender never earn larger payoffs than she/he does nothing no matter how the threshold is set. Thus, "attack, not monitor" is the NE and $x^* = x''$. The sufficiency is shown.

(2) Necessity: under the strategy combination "attack, monitor," the defender and attacker will select $\underline{y}$ and x' as the optimal detection threshold and attack intensity from (4) and (2). If $-q(x'', \underline{y})s(x'') - c_2 \leq -s(x'')$, then similar to (5), there is

$$\begin{aligned} &-q(x', y)s(x') - c_2 \\ &\quad \leq -q(x', \underline{y})s(x') - c_2 \\ &\quad \leq -s(x'). \end{aligned} \tag{6}$$

This means the defender never earns larger payoffs than she/he does nothing, which is a contradiction to the premise that "attack, monitor" is the NE. Thus, the necessity is shown.

Sufficiency: the attacker always selects "attack" from the definitions of π_2 and π_3. If the attacker selects $x^* = x'$, since $-q(x', \underline{y})s(x') - c_2 > -s(x')$, the defender will select "monitor" to obtain larger payoffs than "not monitor" and the optimal detection threshold is $\underline{y}$ from (4). Meanwhile, when the defender selects "monitor" and $y^* = \underline{y}$, from (2), the attack will select "attack" and $x^* = x'$ to earn the maximum positive payoff. Thus, the sufficiency is shown.

Based on Theorem 4, the uniqueness of the NE for Scenario L.4 can also be concluded.

Corollary 1. *In Scenario L.4,*

(1) If and only if $-q(x'', \underline{y})s(x'') - c_2 \leq -s(x'')$ *and* $-q(x', \underline{y})s(x') - c_2 \leq -s(x')$, *"attack, not monitor" is the unique NE and* $x^* = x''$

(2) If and only if $-q(x'', \underline{y})s(x'') - c_2 > -s(x'')$ *and* $-q(x', \underline{y})s(x') - c_2 > -s(x')$, *"attack, monitor" is the unique NE and* $x^* = x'$ *and* $y^* = \underline{y}$

Proof. From Theorem 4, "attack, not monitor" and "attack, monitor" are the only two possible NEs. Clearly, "attack, not monitor" is the unique NE if an extra condition holds, i.e., $-q(x'', \underline{y})s(x'') - c_2 \leq -s(x'')$. Then, "attack, monitor" will not be the NE. Similarly, "attack, monitor" is the unique NE if the extra condition $-q(x', \underline{y})s(x') - c_2 > -s(x')$ holds. Then, "attack, not monitor" is not the NE. Therefore, Corollary 1 can be concluded. □

Scenario L.5. $\pi_1 \neq \varnothing$, $\pi_2 \neq \varnothing$, $\pi_3 \neq \varnothing$, and $\pi_4 = \varnothing$.

Different from Scenario L.4, there exists $x \in [\underline{x}, \overline{x}]$ belonging to π_1 such that $s(x) - c_1 - u(x) \leq 0$. Since the attacker can always find an x such that she/he earns a positive payoff, the strategy combination "not attack, not monitor"

cannot be the NE in this scenario. The main results about the NE in this scenario can be formally stated in the following theorem.

Theorem 5. *In Scenario L.5,*

(1) If and only if $-q(x'', \underline{y})s(x'') - c_2 \leq -s(x'')$, *"attack, not monitor" is the NE and* $x^* = x''$

(2) If and only if $q(x', \underline{y})s(x') - c_1 - u(x') > 0$ *and* $-q(x', \underline{y})s(x') - c_2 > -s(x')$, *"attack, monitor" is the NE and* $x^* = x'$ *and* $y^* = \underline{y}$

(3) If and only if $-q(x'', \underline{y})s(x'') - c_2 \leq -s(x'')$ *and* $\{q(x', \underline{y})s(x') - c_1 - u(x') \leq 0$ *or* $-q(x', \underline{y})s(x') - c_2 \leq -s(x')\}$, *"attack, not monitor" is the unique NE and* $x^* = x''$

(4) If and only if $q(x', \underline{y})s(x') - c_1 - u(x') > 0$, $-q(x', \underline{y})s(x') - c_2 > -s(x')$, *and* $-q(x'', \underline{y})s(x'') - c_2 > -s(x'')$, *"attack, monitor" is the unique NE and* $x^* = x'$ *and* $y^* = \underline{y}$

Proof

(1) The proof is similar to that of (1) in Theorem 4 and is omitted here.

(2) Different from Scenario L.4, there exists $x \in [\underline{x}, \overline{x}]$ belonging to π_1 such that

$$q(x, y)s(x) - c_1 - u(x) < s(x) - c_1 - u(x) \leq 0, \quad (7)$$

as $q(x, y) < 1$. Thus, compared to (2) in Theorem 4, an extra condition $q(x', \underline{y})s(x') - c_1 - u(x') > 0$ needs be added to ensure that "attack, monitor" still be the NE. The remaining proof is similar to that of (2) in Theorem 4 and is omitted here.

(3) and (4) By following similar analysis in the proof of Corollary 1, the uniqueness of the NE in this case can also be concluded.

In contrast to previous scenarios, π_4 and at least one subset of π_1, π_2, and π_3 are nonempty in Scenario L.6 as described below. □

Scenario L.6. $\pi_1 \cup \pi_2 \cup \pi_3 \neq \varnothing$, and $\pi_4 \neq \varnothing$.

From (4) in Theorem 1, there is no NE if only $\pi_4 \neq \varnothing$. Besides, if $\pi_4 = \varnothing$ is replaced by $\pi_4 \neq \varnothing$ for (1)–(3) in Scenario L.1 and Scenarios L.2–L.5, the NEs will never belong to π_4. This is because all the strategy combinations driven by x and y within π_4 are inconsistent with the obtained NE in Theorems 1–5. Hence, (x^*, y^*) of the NE for Scenario L.6 will belong to π_1, π_2, or π_3. Moreover, the conditions for the derived NEs in Theorems 1–5 are still necessary. Therefore, to analyze the NE in Scenario L.6, we only need to verify whether the results in Theorems 1–5 are still correct if the subset π_4 is changed to be nonempty. The following conclusions will be shown.

Theorem 6. *In Scenario L.6, the NE for the game as described in Table 1 is derived as follows:*

(1) If $\pi_1 \neq \varnothing$ *and* $\pi_2 = \pi_3 = \varnothing$, *no NE exists*

(2) If $\{\pi_2 \neq \varnothing, \pi_1 = \pi_3 = \varnothing\}$ *or* $\{\pi_1 \neq \varnothing, \pi_2 \neq \varnothing, \pi_3 = \varnothing\}$, *the results in (1) in Theorem 4 hold true and "attack, not monitor" is the unique NE*

(3) If $\{\pi_3 \neq \varnothing, \pi_1 = \pi_2 = \varnothing\}$ *or* $\{\pi_1 \neq \varnothing, \pi_3 \neq \varnothing, \pi_2 = \varnothing\}$, *the results in Theorem 3 hold true*

(4) If $\{\pi_2 \neq \varnothing, \pi_3 \neq \varnothing, \pi_1 = \varnothing\}$ *or* $\{\pi_1 \neq \varnothing, \pi_2 \neq \varnothing, \pi_3 \neq \varnothing\}$, *the results in Theorem 5 hold true*

Proof

(1) As there exists an x such that the payoff of the attacker $s(x) - c_1 - u(x)$ is positive, "not attack, not monitor" is no longer the NE if $\pi_4 = \varnothing$ is replaced by $\pi_4 \neq \varnothing$ for (1) in Scenario L.1, i.e., $\pi_1 \neq \varnothing$, $\pi_2 = \pi_3 = \varnothing$, and $\pi_4 \neq \varnothing$. It can be easily shown that other strategy combinations cannot be the NE either.

(2) If $\pi_4 = \varnothing$ is replaced by $\pi_4 \neq \varnothing$ for (2) in Scenario L.1, there exists feasible x and y such that $-q(x, y)s(x) - c_2 > -s(x)$. Thus, an extra condition $-q(x'', \underline{y})s(x'') - c_2 \leq -s(x'')$ is required with comparison to (2) in Theorem 1 to ensure that "attack, not monitor" still be the NE. If $\pi_4 = \varnothing$ is replaced by $\pi_4 \neq \varnothing$ for Scenario L.2, by following similar analysis in the proofs of Theorem 2 and (1) in Theorem 4, we can show that the results in (1) in Theorem 4 are true.

(3) When $\pi_3 \neq \varnothing$, $\pi_1 = \pi_2 = \varnothing$, and $\pi_4 \neq \varnothing$, there exist feasible x and y such that $q(x, y)s(x) - c_1 - u(x) < 0$. Thus, an extra condition $q(x', \underline{y})s(x') - c_1 - u(x') > 0$ is required with comparison to (3) in Theorem 1 to ensure that "attack, monitor" still be the NE. When $\pi_1 \neq \varnothing$, $\pi_3 \neq \varnothing$, $\pi_2 = \varnothing$, and $\pi_4 \neq \varnothing$, based on the proof of Theorem 3 and the definitions of π_3 and π_4, it can be shown that the results of Theorem 3 are still true.

(4) Firstly, if π_4 is changed to be nonempty in Scenario L.4, x and y belonging to π_4 will have no influence on the results of (1) in Theorem 4. As the results of (1) in Theorem 5 are the same as that of (1) in Theorem 4, (1) in Theorem 5 holds true in this case. Besides, an extra condition $q(x', \underline{y})s(x') - c_1 - u(x') > 0$ is required with comparison to (2) in Theorem 4 to ensure "attack, monitor" be the NE since there exist x and y such that $q(x, y)s(x) - c_1 - u(x) < 0$ from the definition of π_4. Thus, the results in (2) in Theorem 5 are true. The uniqueness of the NE can also be verified from (3) and (4) in Theorem 5. Secondly, if all the subsets are nonempty, i.e., $\pi_1 \neq \varnothing$, $\pi_2 \neq \varnothing$, $\pi_3 \neq \varnothing$, and $\pi_4 \neq \varnothing$, it can be easily shown that the feasible values of x and y belonging to π_4 have no influence on the results of Theorem 5. □

Remark 3. It can be seen from (3) in Theorem 1, Theorem 3, (2) in Theorem 4, and (2) in Theorem 5 that once the defender decides to monitor in (3) in Scenario L.1, Scenario L.3, (2) in Scenario L.4, (2) in Scenario L.5, and (4) and (5) in

Scenario L.6, she/he will always select $\underline{y}$ as the optimal threshold y^*.

Remark 4. In this paper, we assume that the attackers are completely rational, while this assumption may not be satisfied in some scenarios. However, based on our method, we present an optimal defense strategy for the worst case. That is, we can guarantee that the maximum damage in the worst case can be minimized by our method.

4. Simulation Studies

In this section, simulation results are provided to validate the theoretical results as presented above. In A-IDS, a profile is generally selected to cause distinctions between normal and abnormal states. Such a profile is normally described by a random variable in many cases. Here, we assume it follows a Gaussian distribution with zero mean under normal states. Similar assumptions can be seen in many intrusion detection application areas such as network traffic detection and Kalman filtering-based anomaly detection. Let the intensity of the attack be denoted as x. Other parameters in simulation are chosen as $\underline{x} = \underline{y} = 0.1$, $\overline{x} = \overline{y} = 2$, $s = 0.5x$, $u = 0.1x$, and $c_3 = 0.2$. The false alarm rate and missing alarm rate can be expressed by

$$p = \left(\int_{y}^{\infty} e^{-z^2/8} \mathrm{d}z \right) / 2\sqrt{2\pi},$$
$$q = \left(\int_{-\infty}^{y} e^{-(z-x)^2/8} \mathrm{d}z \right) / \sqrt{2\pi}, \tag{8}$$

respectively. Parameters c_1 and c_2 are used to represent the costs of the attacker and the defender, respectively.

Case 1. We first select $c_1 \in [0, 0.2]$ and $c_2 = 0.2$. Then, it can be calculated by (1) that

(a) If $c_1 \in [0, 0.04]$, there are $\pi_2 \neq \varnothing$, $\pi_3 \neq \varnothing$, and $\pi_1 = \pi_4 = \varnothing$, which corresponds to Scenario L.4

(b) If $c_1 \in [0.04, 0.08]$, there are $\pi_1 \neq \varnothing$, $\pi_2 \neq \varnothing$, $\pi_3 \neq \varnothing$, and $\pi_4 = \varnothing$, which corresponds to Scenario L.5

(c) If $c_1 \in [0.08, 0.2]$, all the four subsets are nonempty, which corresponds to Scenario L.6

Then, it can be checked whether the inequality conditions in Theorems 4 and 5 and (4) in Theorem 6 are satisfied for the above three scenarios, as given in Table 3. 'IC 4.1', 'IC 4.2', 'IC 5.1', and 'IC 5.2' refer to the inequality conditions in (1) and (2) in Theorem 4 and (1) and (2) Theorem 5, respectively. It is worth noting that the inequality conditions in (4) in Theorem 6 are the same as those in Theorem 5. From the theoretical analysis given in Section 2, the following conclusions on the NEs can be drawn:

(a) Based on (2) in Theorem 4, "attack, monitor" is the unique NE if $c_1 \in [0, 0.04]$ and $c_2 = 0.2$.

(b) Based on (2) in Theorem 5, "attack, monitor" is the unique NE if $c_1 \in [0.04, 0.08]$ and $c_2 = 0.2$.

(c) Based on (4) in Theorem 6 and (2) in Theorem 5, "attack, monitor" is still the unique NE if $c_1 \in [0.08, 0.2]$ and $c_2 = 0.2$. However, no NE exists if $c_1 \in (0.12, 0.2]$, $c_2 = 0.2$. This result can be verified by observing the payoff of the attacker (U_A) with respect to c_1, as shown in Figure 1. U_A decreases as c_1 increases. Besides, U_A will approach zero when c_1 tends to 0.12, which indicates that the NE is broken.

Case 2. In this case, we fix c_1 as $c_1 = 0.1$, while let c_2 vary within the interval $[0, 0.2]$. It can be calculated that

(a) If $c_2 \in [0, 0.04]$, there are $\pi_1 \neq \varnothing$, $\pi_3 \neq \varnothing$, $\pi_4 \neq \varnothing$, and $\pi_2 = \varnothing$, which corresponds to Scenario L.6

(b) If $c_2 \in [0.04, 0.2]$, all the four subsets are nonempty, which also corresponds to Scenario L.6

Similarly, Table 4 is given to show whether the inequality conditions in Theorems 3 and 5 are satisfied, where 'IC 3' refers to the inequality condition in Theorem 3. Then, the following conclusions on the NEs can be drawn:

(a) Based on Theorem 3 and (3) in Theorem 6, "attack, monitor" is the unique NE if $c_1 = 0.1$ and $c_2 \in [0, 0.04]$

(b) Based on (2) in Theorem 5 and (4) in Theorem 6, "attack, monitor" is the unique NE if $c_1 = 0.1$ and $c_2 \in (0.04, 0.2]$

Therefore, "attack, monitor" is always the unique NE if $c_1 = 0.1$, $c_2 \in [0, 0.2]$. Besides, from Theorem 3 and (2) Theorem 5, it can be calculated that the payoff of the attacker (U_A) is equal to 0.024 if $c_2 \in [0, 0.2]$. It indicates that the attacker has the motivation to launch the attack. The performance of the defender's payoff (U_D) with respect to c_2 is shown in Figure 2. Clearly, the defender loses some security asset as $U_D < 0$. Moreover, the lost security asset will increase as the defense cost c_2 increases.

At last, we make some comparisons with the existing methods in [7–15], where attack intensity and detection threshold are scarcely considered and majority of them assume that the false and missing alarm rates, and the game model of detection problem can be modelled as Table 5.

It can be seen that, without considering the attack intensity and detection threshold, the payoffs of the game model will be reduced to be constant and the Nash equilibrium analysis can be easily done. From the definition of the Nash equilibrium, it can be calculated that if $q + c_2 > 1$, (Attack, Monitor) will be the unique NE. Though the existing analysis methods in [7–15] can determine the optimal action strategies, while our results can further determine the optimal explicit attack intensity and detection threshold, different results can be obtained. First, the existing work considers only the strategy do or not do; thus, the one-stage game model, as expressed in Table 3, is established to help analyze the optimal actions, while we further consider the attack intensity and detection threshold in the game model, as these two parameters are two key strategies used for the defender and the attacker. Moreover, we establish a more detailed two-stage game model to consider both the action do or not do and the attack intensity

Table 3: The results showing whether the inequality conditions in Theorems 4 and 5 and (4) in Theorem 6 are satisfied when $c_1 \in [0, 0.2]$ and $c_2 = 0.2$.

$c_1 \in [0, 0.04]$ Scenario L.4		$c_1 \in [0.04, 0.08]$ Scenario L.5		$c_1 \in [0.08, 0.12]$ Scenario L.6		$c_1 \in (0.12, 0.2]$ Scenario L.6	
IC 4.1	×	IC 5.1	×	IC 5.1	×	IC 5.1	×
IC 4.2	√	IC 5.2	√	IC 5.2	√	IC 5.2	×

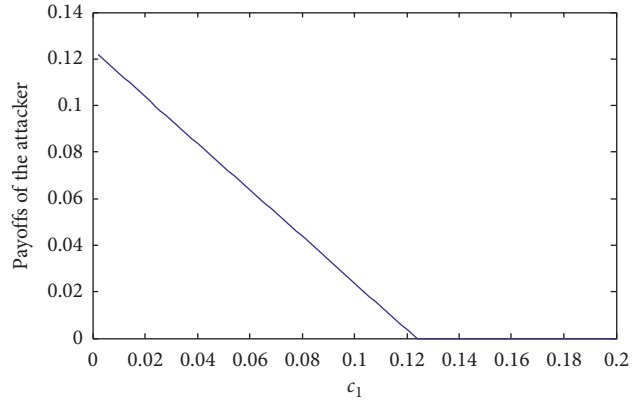

Figure 1: Payoff of the attacker U_A with respect to c_1 if c_2 is fixed as $c_2 = 0.2$.

Table 4: The results about the inequality conditions in Theorems 3 and 5 with $c_1 = 0.1$ and $c_2 \in [0, 0.2]$.

$c_2 \in [0, 0.04]$ Scenario L.4		$c_2 \in (0.04, 0.2]$ Scenario L.5	
IC 3	√	IC 5.1	×
		IC 5.2	√

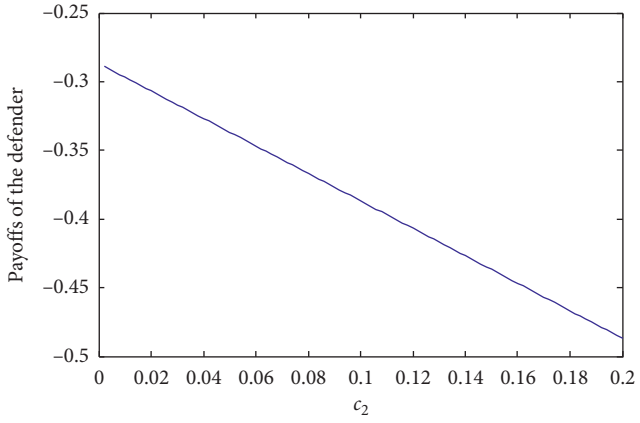

Figure 2: Payoff of the defender U_D with respect to c_2 if c_1 is fixed as $c_1 = 0.1$.

Table 5: Strategic form of the game in existing work.

	Monitor	Not monitor
Attack	$U_A = qW - c_1W$ $U_D = -qW - c_2W$	$U_A = W - c_1W$ $U_D = -W$
Not attack	$U_A = 0$ $U_D = -pc_3W - c_2W$	$U_A = 0$ $U_D = 0$

and detection threshold. Based on the experimental results, we can see that the attack intensity and detection threshold play an important role in the determination of the Nash equilibrium. Intuitively, for the game in Table 3, the NE are completely determined by the parameter x and y; however, this conclusion seems not to make sense as the false alarm rate and other parameters have no any effect on the Nash equilibrium. Alternatively, for our game model, we can see that all parameters will jointly determine the Nash equilibriumthus, our analysis results are more realistic. In practical, the false and missing alarm rates are not constant, as the attacks are always dynamically changing. In A-IDS methods, the false and missing alarm rates are commonly determined by the attack intensity and detection threshold. Our method just considers this real scenario and establishes a more explicit game model, based on which the optimal strategies are completely determined.

5. Conclusion

For anomaly-based intrusion detection system, we present a game theoretical analysis method to provide the optimal strategies. We first establish a more realistic game model by considering the attack intensity and detection threshold as two strategies for the players. The necessary and sufficient conditions, for which strategies are the Nash equilibriums, are presented. Simulation studies are provided to validate our theoretical results. The results provide a new method to determine the detection threshold in the security defense. In the future, some more research work could be considered, for example, the game theoretical analysis method for

specific scenarios such as Internet of Things and DoS/DDoS attacks. Besides, dynamic game analysis is also an interesting topic for dynamic security confrontation process, for example, Stackelberg game analysis can be adopted to solve the sequential problem of the attack and defense actions.

Data Availability

The manuscripts of game theory algorithm in this article are from the databases of Cambridge University and Columbia University. Copies of these data can be obtained from https://dl.acm.org/doi/book/10.5555/1951874 and https://doi.org/10.1016/j.ins.2018.04.051.

Conflicts of Interest

The authors declared that they have no conflicts of interest.

Acknowledgments

This work was supported by the Basic Scientific Research Projects of National Defense Science, Technology and Industry Technology under Grant no. JSZL2017601C-1 and in part by the National Natural Science Foundation of China under Grant nos. 61897069 and 61831003, National Key Research and Development Program of China under Grant no. 2017YFB0801903, and National Key Program for Basic Research of China under Grant no. 2017-JSJQ-ZD-043.

References

[1] H. J. Liao, C. H. R. Lin, Y. C. Lin, and K. Y. Tung, "Intrusion detection system: a comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[2] P. G. Teodoro, J. D. Verdejo, G. M. Fernandez, and E. Vazquez, "Anomaly-based network intrusion detection: techniques, systems and challenges," *Computers & Security*, vol. 28, no. 1-2, pp. 18–28, 2009.

[3] M. Manshaei, Q. Zhu, T. Alpcan, T. Basar, and J. P. Hubaux, "Game theory meets network security and privacy," *ACM Computing Surveys*, vol. 45, no. 3, pp. 1–39, 2013.

[4] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A survey of game theory as applied to network security," in *Proceedings of the 43rd Hawaii International Conference on System Sciences*, IEEE, Honolulu, HI, USA, January 2010.

[5] X. Liang and Y. Xiao, "Game theory for network security," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 472–486, 2013.

[6] C. Manikopoulos and S. Papavassiliou, "Network intrusion and fault detection: a statistical anomaly approach," *IEEE Communications Magazine*, vol. 40, no. 10, pp. 76–82, 2002.

[7] T. Alpcan and T. Basar, *Network Security: A Decision and Game-Theoretic Approach*, Cambridge University Press, Cambridge, UK, 2011.

[8] L. Chen and J. Leneutre, "A game theoretical framework on intrusion detection in heterogeneous networks," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 2, pp. 165–178, 2009.

[9] Z. Ismail and J. Leneutre, "A game theoretical analysis of data confidentiality attacks on smart-grid AMI," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 7, pp. 1486–1499, 2014.

[10] Q. Zhu, C. Fung, R. Boutaba, and T. Basar, "GUIDEX: A game-theoretic incentive-based mechanism for intrusion detection networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 11, pp. 2220–2230, 2012.

[11] H. Wu, W. Wang, C. Wen, and Z. Li, "Game theoretical security detection strategy for networked systems," *Information Sciences*, vol. 453, pp. 346–363, 2018.

[12] Y. Liu, C. Comaniciu, and H. Man, "A Bayesian game approach for intrusion detection in wireless ad hoc networks," *ACM International Conference Proceeding Series*, vol. 199, 2006.

[13] K. C. Nguyen, T. Alpcan, and T. Basar, "Security games with incomplete information," in *Proceedings of the of 2009 IEEE International Conference on Communications*, IEEE, Dresden, Germany, June 2009.

[14] W. Wang, M. Chatterjee, and K. Kwiat, "Attacker detection game in wireless networks with channel uncertainty," in *Proceedings of the 2010 IEEE International Conference on Communications*, IEEE, Cape Town, South Africa, May 2010.

[15] Y. E. Sagduyu, R. Berry, and A. Ephremides, "MAC games for distributed wireless network security with incomplete information of selfish and malicious user types," in *Proceedings of the 2009 International Conference on Game Theory for Networks*, IEEE, Istanbul, Turkey, May 2009.

[16] A. Bradai and H. Afifi, "Game theoretic framework for reputation-based distributed intrusion detection," in *Proceedings of the 2013 International Conference on Social Computing*, IEEE, Alexandria, VA, USA, September 2013.

[17] W. Wang, M. Chatterjee, K. Kwiat, and Q. Li, "A game theoretic approach to detect and co-exist with malicious nodes in wireless networks," *Computer Networks*, vol. 71, pp. 63–83, 2014.

[18] W. Yu and K. J. R. Liu, "Secure cooperation in autonomous mobile ad-hoc networks under noise and imperfect monitoring: a game-theoretic approach," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 317–330, 2008.

[19] F. Li, Y. Yang, and J. Wu, "Attack and flee: game-theory-based analysis on interactions among nodes in MANETs," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 40, no. 3, pp. 612–622, 2010.

[20] L. Xiao, Y. Chen, W. S. Lin, and K. J. R. Liu, "Indirect reciprocity security game for large-scale wireless networks," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1368–1380, 2012.

[21] H. Moosavi and F. M. Bui, "A game-theoretic framework for robust optimal intrusion detection in wireless sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 9, pp. 1367–1379, 2014.

[22] S. A. Zonouz, H. Khurana, W. H. Sanders, and T. M. Yardley, "RRE: A game-theoretic intrusion response and recovery engine," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 395–406, 2014.

[23] H. Wu and W. Wang, "A game theory based collaborative security detection method for Internet of Things systems," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 6, pp. 1432–1445, 2018.

[24] H. Wu and Z. Wang, "Multi-source fusion-based security detection method for heterogeneous networks," *Computers & Security*, vol. 74, pp. 55–70, 2018.

[25] R. Jin, X. He, and H. Dai, "On the security-privacy tradeoff in collaborative security: a quantitative information flow game perspective," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3273–3286, 2019.
[26] H. Zhang, L. Jiang, S. Huang, J. Wang, and Y. Zhang, "Attack-defense differential game model for network defense strategy selection," *IEEE Access*, vol. 7, pp. 50618–50629, 2018.
[27] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.