Moving Target Defense Decision-Making Method: A Dynamic Markov Differential Game Model

Hengwei Zhang
State Key Laboratory of Mathematical Engineering and Advanced Computing
Zhengzhou, China
wlby_zzmy_henan@163.com

Jinglei Tan
State Key Laboratory of Mathematical Engineering and Advanced Computing
Zhengzhou, China
nxutjl@126.com

Xiaohu Liu
State Key Laboratory of Mathematical Engineering and Advanced Computing
Zhengzhou, China
yudianhappy@126.com

Jindong Wang
State Key Laboratory of Mathematical Engineering and Advanced Computing
Zhengzhou, China
wangjindong_hnxd@126.com

ABSTRACT

Today most of the moving target defense decision-making methods are based on models of a discrete dynamic game. To more accurately study network attack-defense strategies against continuous confrontations, we analyze offensive and defensive behavior from a dynamic perspective. We propose a moving target defense decision-making method based on a model of a dynamic Markov differential game. We implement dynamic analysis and deduction of multi-stage continuous attack and defense confrontations for scenarios of continuous real-time network attack-defense. We take into account the influence of random factors and changes of the network system in the gaming process, combine differential gaming with the Markov decision-making method, and construct models of attack-defense games. We propose a solution for game equilibrium based on an objective function designed according to the total discounted payoff of the offensive and defensive game and the analysis of the characteristics of multi-staged game equilibrium. On this basis an optimal strategy selection method is designed. We apply and verify the game model and the defense strategy selection algorithm by using the moving target defense technique. We conduct simulations to verify the effectiveness and feasibility of the model and algorithm.

CCS CONCEPTS

• CCS → Security and privacy → Network security.

KEYWORDS

Moving target defense, Differential game, Markov decision-making, Defense strategy selection.

ACM Reference format:
Hengwei Zhang, Jinglei Tan, Xiaohu Liu, and Jindong Wang. 2020. Moving Target Defense Decision-Making Method: A Dynamic Markov Differential Game Model. In *6th ACM Workshop on Moving Target Defense (MTD'20), November 11, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3411496.3421222

Corresponding author: Jinglei Tan (nxutjl@126.com)

MTD'20, November 9, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-8085-0/20/11...$15.00
https://doi.org/10.1145/3411496.3421222

1 INTRODUCTION

Network technologies are now widely used in a variety of fields, including politics, economics, military, society, and education. They have propelled human society from an industrial age to an information age [1]. The Internet has brought tremendous progress and convenience to human activities, but at the same time, attacks have also brought on frequent network security incidents and a severe network security situation. Network attack technologies have had unabated development. Due to considerations of development cost and post-maintenance, most of the existing network systems have static, deterministic, and isomorphic architecture designs. The flaws and omissions of this type of design have resulted in inevitable security vulnerabilities of the network system [2]. Although current network defense technologies, such as a firewall, vulnerability scan, and intrusion detection systems, have gained maturity and achieved good defensive effects; most of the existing defense technologies rely on prior knowledge and are passive and lagging. These features provide the attacker an asymmetric advantage in time and cost. Therefore, it is very difficult for the existing passive defense technology to effectively cope with complex and versatile cyber-attacks with limited resources; the defenders are always at a disadvantage for being a "sitting duck".

In order to change this asymmetric passive cyber defense situation, active defense technologies have now appeared. The Moving Target Defense (MTD) is a typical active defense technology. Its core idea is to analyze, create, and deploy diverse and continuously changing mechanisms and strategies to enhance the defense capability of the target system by increasing the complexity and overhead of the attack and reducing the vulnerability exposure and the probability for exploitation [3]. In attack-defense confrontations the key to achieve the goal is choosing the best strategy for action. This has become a hot topic in recent years. Analysis of the confrontation characteristics of MTD reveals that both the offense side and the defense side often have conflicting goals, a non-cooperative relationship, and strategic dependence. These characteristics happen to coincide with the basic features of game theory [4]. For this reason theoretical studies of the MTD attack-defense process in game theory and solutions for the optimal defense strategy are of particular value for theory and engineering.

At present some results have been achieved by using game theory to study the selection of MTD strategies. In Ref. [5] the dynamic defense process is abstracted into a dynamic game of two players with incomplete information. Both the static attack scenario and the adaptive attack scenario are considered, and platform diversity is used as an indicator for choosing the optimal defense strategy. In Ref. [6] an empirical game method was proposed to analyze the dynamic confrontation process between the attacker behavior and the MTD strategy. Reference [7] considered the network stack configuration transfer in MTD and proposed a repeated Bayesian game model to abstract the optimal transfer strategy as a balance between transfer strategy cost and transfer strategy generation efficiency. In [8] a multi-stage game model with an information feedback mechanism was designed for MTD. An algorithm for an optimum defense strategy solution was designed using stochastic dynamics. Analysis of the attack surface transfer process in a multi-staged target system was based on a mixed strategy Nash equilibrium. In [9] a MTD model based on a multi-stage signal game was proposed, in which the defender was regarded as the signal sender, and the attacker was regarded as the signal receiver. The defender may design a signal sending mechanism to affect the attack strategy and to design an algorithm for selecting an optimum defense strategy using the uncertainty created by MTD. The above results were all based on the model for the incomplete information dynamic game that studied the time-discrete attack and defense confrontation process. However, as the network attack and defense process becomes more dynamic, time becomes more continuous, and the action frequency increases. It becomes more difficult for traditional game models to meet the need of MTD in the real-time confrontation environment.

A differential game is a theoretical method for studying conflicts under the condition of continuous time variation [10]. It has been applied in the fields of network defense, security threat warning, and emergency decision-making. In [11] a differential game model of attack-defense is constructed to analyze the dynamic, continuous attack-defense process through simulations and to study the changes of general behavior. However, the lack of specific algorithms for solving the saddle point strategy and the lack of processes for analyzing a game equilibrium has reduced the application value. In [12] a model for network defense based on a differential game was proposed, which achieved the analysis and deduction of real-time network attack-defense. On the basis of the algorithm for game equilibrium, an algorithm for selecting a real-time optimal defense strategy was designed. In [13] a model of a continuous differential game was based on the uncertainty and dynamics of incidents from the perspective of multi-agent competition, and taking into account both optimal risk control and self-reward maximization. The model combined dynamic game theory and optimization control theory to design an algorithm for selecting the optimum strategy for decision makers and regulators. In [14] regional economic cooperation behavior is abstracted into differential games, and game models for weak-weak confrontation, strong-weak master-slave, and strong-strong collaboration were constructed. In addition the relationships between individual optimal strategy, maximum benefit, and overall income were analyzed. The above researches have successfully carried out modeling research on network defense, risk management, and economic decision-making under continuous time conditions.

Since the actual confrontation process of MTD is dynamic and changes, and the state of the target system is affected by such random factors as network environment and changes of the attack-defense strategy, we introduce the Markov decision process to characterize the real-time stochastic jumps of the target state. Based on the above analysis, we construct a model of a dynamic Markov differential game to describe the real-time network attack-defense process. In view of the payoff attenuation in the offensive and defensive process, a discount factor is introduced for the game payoff at different stages, and the total discounted payoff is used as the objective function. On this basis we analyzed and solved the equilibrium path, designed the optimal defense selection algorithm, and implemented the simulations based on MTD technology to verify the feasibility and effectiveness of the model. When compared to prior work, the model proposed here can guide and inspire theoretical development of defense decision-making for real-time continuous network attack and defense scenarios.

2 Construction of the model for the Dynamic Markov Differential Game

2.1 Analysis of offensive and defensive game processes

A Differential Game is a theoretical method for describing the continuous control process of a confrontation under real-time changes. It is suitable for studying real-time control optimization for all parties in a conflict. Today the attack and defense confrontation in the field of network security is becoming increasingly acute and is no longer a time-discrete, round-interactive attack and defense process. The network attack and defense process is characterized by high frequency conflict action, continuous confrontation time, and control decision-making in real time. Conventional game theory can no longer meet the needs in reality, but the method for the differential game has distinct advantages in the study of network attack and defense problem.

When using classical differential game theory in the study of the offensive and defensive game process, external interference conditions are usually not considered. It is assumed that the optimal strategy (also called optimal control) of both sides of the game is a control function with time as the independent variable, and that the game payoff and system state of both parties are deterministic and predictable. However, in the actual network system operation process and the attack and defense confrontation process, the presence of various inevitable interference and random factors such as user demand changes, system parameter changes, and accidental disturbances makes the state changes of the network system dynamic and uncertain. Therefore, the study of offense and defense games must include the influence of random interference factors. Based on the above analysis, we use the model for the differential game to analyze the MTD confrontation process under continuous time conditions. For random interference factors in the attack and defense game process and potential stochastic jumps of the network system state, we use the Markov decision process in our analysis. By combining the differential game and Markov decision process, we constructed a model of a dynamic Markov differential game to address the problem of optimal defense strategy selection.

A schematic diagram of the offensive and defensive game process is shown in Figure 1. It is a multi-stage, multi-state process where the state changes dynamically with time. In a given time period both offensive and defensive sides start from the initial state, continuously make decisions in the dynamic confrontation, and evolve toward the subsequent state. With the passage of time the systems are, on the one hand, affected by the offensive and defensive game behavior, and on the other hand, affected by the

system environment and potential changes of the game elements. With a probability of η the network system jumps from one state to another, and the system enters in a new phase of the game in a new state. From a global perspective of the complete offensive and defensive game process, the system is in the dynamic process of jumping from "attack and defense game - state to - offensive and defensive game." Based on the above analysis, we construct a model of a dynamic Markov differential game, introduce a discount factor μ, and quantitatively calculate the expected total payoff for both offensive and defensive sides from the initial stage to the final stage. Using this as an objective function for both sides, we find a solution for game equilibrium and selected a defense strategy.

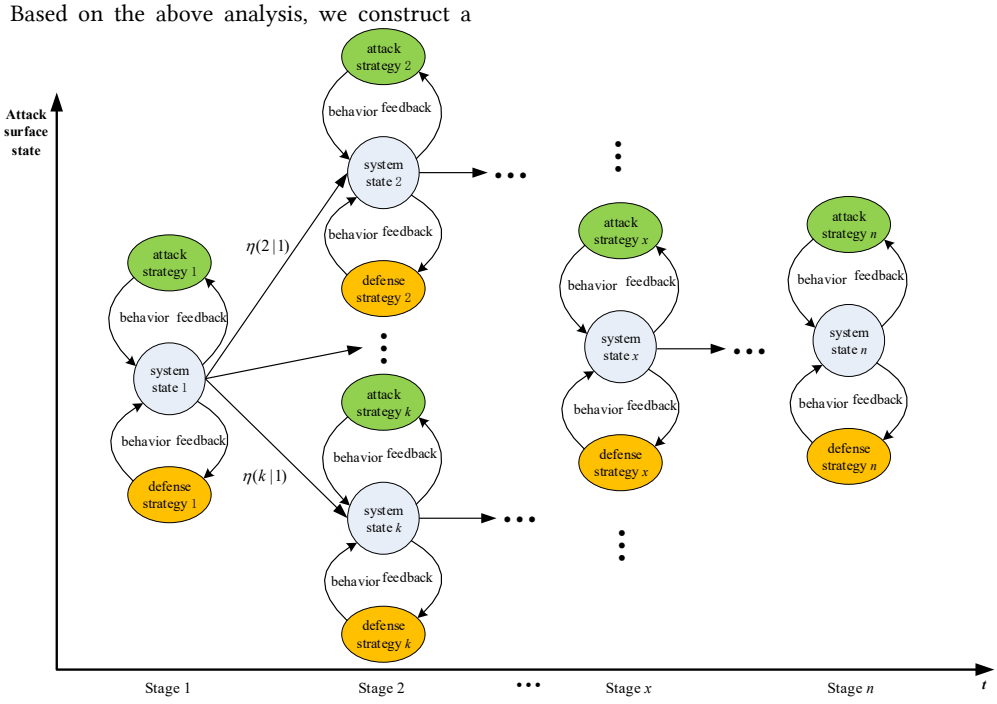


Figure 1: Moving target defense Markov differential game process

2.1 Game Model Construction

Definition 1 The Moving Target Defense Dynamic Markov Differential Game Model (TMDGM) can be expressed as a decuple: $TMDGM = (N, \Theta, T, t, B, S, \eta, x, \mu, U)$

(1) $N = (N_a, N_d)$ is a collection of game participants: N_a attackers and N_d defenders.

(2) $\Theta = (\Theta_a, \Theta_d)$ is the type of space of the attacker and defender, Θ_a indicates the type of attackers and Θ_d indicates the type of defenders.

(3) T indicates the total number of stages of the differential game, $G(k)$ indicates the k-th stage of the game process, and $1 \le k \le T, k \in \mathrm{N}$.

When the game process is in the $G(k)$ stage, the attacker employs the means of active detection and public information gathering to obtain attack surface [15, 16, 17] information of the current target system and accordingly determines the attack strategy and builds the attack chain. The defender then selects the corresponding defense strategy to destroy the integrity of the attack chain and implements dynamic defense. The two sides then enter an offensive and defensive game phase in a relatively stable system environment. When the operating environment of the target system is stochastically disturbed by interference and accidental factors, the system state stochastically jumps to the next state, and the game process enters the $G(k+1)$ phase.

(4) t indicates the time in the differential game, $t \in [0, t_T]$, t_0^k indicates the initial time of the k-th stage, t_k indicates the end time of the k-th stage.

(5) $B = \{(a(t), d(t)) \mid a(t) \in A, d(t) \in D\}$ is the strategy space of the game, A and D are respectively the set of attack and defense strategies. Denote $a(t)$ and $d(t)$ respectively as the mixed strategy of the attacker and the defender at time t.

(6) Let $S = \{S(k) \mid k = 1, 2, ..., T\}$ represent a set of security states of the network system. For ease of analysis consider only the initial and steady state in the phase, denoted respectively as S_0^k and S_k.

(7) η is the state transfer probability; $\eta_{ij} = \eta(S_0^j \mid S_i)$ indicates the probability that the network system will enter the initial state of the next phase from the stable state of the previous phase under the influence of stochastic factors.

(8) $x(t) = (\alpha(t), \beta(t), \chi(t), \delta(t))$ is the rate of the attack surface resource used by the attacker in the network system. It is the ratio of the attack surface resources AS (including DAS, SAS, NAS, PAS) that the attacker detects or exploits at time t to the total attack surface resources of the network system.

The attacker builds an attack chain based on the public defense information and the actively collected intelligence information. It is assumed that the attack surface resource rate $x(t)$ obeys the classical Logistic process, that is, $\frac{dx}{dt} = px(1 - x / x_m)$ where p is

the intrinsic growth rate and x_m is the critical value of the network system attack surface resource rate.

(9) $\mu(t) = \exp(-\theta(t - t_0^k))$ is the discount factor at time *t*, indicating the discount rate of the current game stage compared to the initial stage.

(10) $U^k = (U_A^k, U_D^k)$ indicates the set of payoff functions of the attack and defense parties in the *k*-th stage.

In [4,7] the statistical mean is used to define the attack return parameter $r = [r_1, r_2, r_3, r_4]$ and the defense return parameter $w = [w_1, w_2, w_3, w_4]$. Here r and w are respectively the degree of impact of the attack surface resources used by the attacker and the defender on the offensive and defensive payoff. In the meantime the attack strategy cost parameter π_A and the defense strategy cost parameter π_D are given.

In summary we obtained the attack return $R_A(t)$ and defense return $R_D(t)$, and the attack cost c_A and defense cost c_D at time *t* as given by Equations (1), (2), (3), and (4):

$$R_A(t) = r \times x(t) = r_1\alpha(t) + r_2\beta(t) + r_3\chi(t) + r_4\delta(t) \tag{1}$$

$$\begin{aligned} R_D(t) &= w \times (AS(t) - x(t)) = w \times \bar{x}(t) \\ &= w_1\bar{\alpha}(t) + w_2\bar{\beta}(t) + w_3\bar{\chi}(t) + w_4\bar{\delta}(t) \end{aligned} \tag{2}$$

$$c_A = \pi_A a^2(\alpha(t) + \beta(t) + \chi(t) + \delta(t)) \tag{3}$$

$$c_D = \pi_D d^2(\bar{\alpha}(t) + \bar{\beta}(t) + \bar{\chi}(t) + \bar{\delta}(t)) \tag{4}$$

For the *k*-th stage attack and defense game the attacker's payoff function U_A^k and the defender's payoff function U_D^k are given by Equations (5) and (6):

$$\begin{aligned} U_A^k &= \int_{t_{k-1}}^{t_k} L_A(t,x,a,d)dt \\ &= \int_{t_{k-1}}^{t_k} \begin{bmatrix} r_1\alpha(t) + r_2\beta(t) + r_3\chi(t) + r_4\delta(t) \\ -\pi_A a^2(\alpha(t) + \beta(t) + \chi(t) + \delta(t)) \end{bmatrix} dt \end{aligned} \tag{5}$$

$$\begin{aligned} U_D^k &= \int_{t_{k-1}}^{t_k} L_D(t,x,a,d)dt \\ &= \int_{t_{k-1}}^{t_k} \begin{bmatrix} w_1\bar{\alpha}(t) + w_2\bar{\beta}(t) + w_3\bar{\chi}(t) + w_4\bar{\delta}(t) \\ -\pi_D d^2(\bar{\alpha}(t) + \bar{\beta}(t) + \bar{\chi}(t) + \bar{\delta}(t)) \end{bmatrix} dt \end{aligned} \tag{6}$$

The design of the objective function J is to characterize the overall payoff of the offensive and defensive game and to judge the merits of both offensive and defensive strategies. Commonly used objective functions have a discount expectation function and an average return function. Since the network attack and defense process is continuous in real time and is subject to random factors, it may cause random jumps of the system state at different stages of the game. Based on the rational person hypothesis, both offensive and defensive sides always aim at a strategy that maximizes their own payoffs. At the same time considering that random factors can affect the offensive and defensive payoffs, we introduce the objective function as Equation (7) with a discount factor to describe the game payoffs:

$$\begin{cases} J_A^k(S_0^k, S_k) = U_A^k(S_0^k, S_k) + \sum_{k<h\le T} \eta(S_0^h \mid S_k) \int_{G(h)} \mu(t) J_A^h(S_0^h, S_h)dt \\ J_D^k(S_0^k, S_k) = U_D^k(S_0^k, S_k) + \sum_{k<h\le T} \eta(S_0^h \mid S_k) \int_{G(h)} \mu(t) J_D^h(S_0^h, S_h)dt \end{cases} \tag{7}$$

3 The Solution for the Game Equilibrium and Design the algorithm for selecting the Defense Strategy

3.1 Game Equilibrium Analysis

Both the attacker and the defender hope to maximize the attack-defense payoffs under real-time continuous time conditions. When investigating the game stage *G*(*k*), and according to the differential game saddle point equilibrium theory [10], if $(a_k(t)^*, d_k(t)^*)$ is the optimal control strategy of the *k*-th stage in the game, then it satisfies Equation (8):

$$\begin{cases} \forall a_k(t), U_A^k(a_k(t)^*, d_k(t)^*) \ge U_A^k(a_k(t), d_k(t)^*) \\ \forall d_k(t), U_D^k(a_k(t)^*, d_k(t)^*) \ge U_D^k(a_k(t)^*, d_k(t)) \end{cases} \tag{8}$$

Since the offense/defense process consists of multiple game stages, each stage will be affected by the attack and defense game behavior of the previous stage. According to Markov decision criteria, each participant must have a Markov optimal response strategy [18]. If $\{(a_k(t)^*, d_k(t)^*) \mid 1 \le k \le T, k \in \mathrm{N}\}$ is a Markov optimal response strategy, then $(a_k(t)^*, d_k(t)^*)$ will make the objective function reach its maximum, that is, the following conditions are met for any phase k as given by Equation (9):

$$\begin{cases} a_k(t)^* = \arg\max J_A^k(S_0^k, S_k) \\ = \arg\max[U_A^k(S_0^k, S_k) + \sum_{k<h\le T} \eta(S_0^h \mid S_k) \int_{G(h)} \mu(t) J_A^h(S_0^h, S_h)\, dt] \\ d_k(t)^* = \arg\max J_D^k(S_0^k, S_k) \\ = \arg\max[U_D^k(S_0^k, S_k) + \sum_{k<h\le T} \eta(S_0^h \mid S_k) \int_{G(h)} \mu(t) J_D^h(S_0^h, S_h)\, dt] \end{cases} \tag{9}$$

The Markov differential game consists of multiple independent and similar differential games. Due to the interference of stochastic effects on the game system, the states jump between the stages. Based on the model definition and the income function definition in Section 2, the differential game in each stage belongs to a finite game. According to the existence theorem of finite game equilibrium [18], the finite game is in Nash equilibrium for mixed strategies. Therefore, the Markov differential game has a mixed strategy Nash equilibrium.

3.2 Determine the solution for Game equilibrium and design the algorithm to select the defense strategy

Combining the analysis in Section 3.1, we constructed the dynamic programming equation to solve the game equilibrium. Based on equations (5), (6), and (7), the objective function of the Markov differential game is given by Equation (10):

$$\begin{cases} \max J_A^k(S_0^k, S_k) = \int_{t_{k-1}}^{t_k} \begin{bmatrix} r_1\alpha(t) + r_2\beta(t) + r_3\chi(t) + r_4\delta(t) \\ -\pi_A a^2(\alpha(t) + \beta(t) + \chi(t) + \delta(t)) \end{bmatrix} dt \\ \qquad + \sum_{k<h\le T} \eta(S_0^h \mid S_k) \int_{G(h)} \mu(t) J_A^h(S_0^h, S_h)dt \\ \max J_D^k(S_0^k, S_k) = \int_{t_{k-1}}^{t_k} \begin{bmatrix} w_1\bar{\alpha}(t) + w_2\bar{\beta}(t) + w_3\bar{\chi}(t) + w_4\bar{\delta}(t) \\ -\pi_D d^2(\bar{\alpha}(t) + \bar{\beta}(t) + \bar{\chi}(t) + \bar{\delta}(t)) \end{bmatrix} dt \\ \qquad + \sum_{k<h\le T} \eta(S_0^h \mid S_k) \int_{G(h)} \mu(t) J_D^h(S_0^h, S_h)dt \end{cases} \tag{10}$$

The constraint is given by Equation (11):

$$s.t.\begin{cases}\dfrac{dx}{dt}=f(t,x(t_1),a(t),d(t))=px(1-x/x_m)\\ x(t)=(\alpha(t),\beta(t),\chi(t),\delta(t))\\ x(0)=x_0\end{cases} \tag{11}$$

In the k-th stage of the game, $a_k(t)$ and $d_k(t)$ represent respectively the control strategies of the defender and the attacker. The optimal control strategy set can be obtained by solving the above dynamic programming problem. Since the model for the differential game generally cannot be solved analytically, a numerical solution that meets the actual accuracy requirements can be obtained within a finite time by using the dynamic programming method and by using MATLAB for the given model characteristics and calculation accuracy.

Based on the above analysis, the algorithm chosen for designing the selection of the optimal defense strategy is as follows.

MTD Optimal Defense Strategy Selection Algorithm

Input: Markov differential game model TMDGM
Output: Optimal defense strategy
BEGIN
1. Initialize game model $TMDGM=(N,\Theta,t,T,B,S,\eta,x,\mu,U)$
2. Construct game strategy set $B=\{(a_k(t),d_k(t))\mid t\in[t_1,t_T],1\le k\le T,k\in \mathrm{N}\}$
3. Construct initial state set for different stages of game and stable state set $\{S_0^1\cdots S_0^k\cdots S_0^T\}$ and $\{S_1\cdots S_k\cdots S_T\}$
4. Initialize state transfer probability $\eta_{ij}=\eta(S_0^j\mid S_i)$ and payoff calculation parameters r,w,π_D,π_A :
5. For (k=1;; k++)
{ //calculate game payoff for various stages
6. Based on Eq. (5) and (6), calculate attacker's and defender's payoff U_A^k for k-th stage
}
7. Based on Eq. (7), calculate attacker's discounted payoff $\sum_{k<h\le T}\eta(S_0^h\mid S_k)\int_{G(h)}\mu(t)J_A^h(S_0^h,S_h)dt$ and defender's discounted payoff $\sum_{k<h\le T}\eta(S_0^h\mid S_k)\int_{G(h)}\mu(t)J_D^h(S_0^h,S_h)dt$
8. Based on Eqs. (10) and (11), solve for optimal strategy set $(a_k^*(t),d_k^*(t))$
9. Output $d_k^*(t)$
END

We compared the model proposed in this paper with the existing research results, as shown in Table 1. This comparative analysis shows that the model of the Markov offense and defense differential game proposed in this paper can transform the sustained network attack and defense into T multiple stages, with each stage lasting for a short period of time. The model is capable of analyzing multi-stage, multi-state, continuous attack and defense process. The objective function based on the total discounted payoff can perform comprehensive calculation in real time on the payoff of multi-staged attack and defense games. The strategy selection algorithm can select the optimal defense strategy for the multi-stage attack and defense process in real time, and has the advantages of clear algorithm steps and easy implementation. Compared to other methods in the literature shown in Table 1, the method proposed here can serve as a better theoretical guide in modeling.

Table 1 Comparative analysis of models and algorithms

Reference	Game process	Game type	Stochastic factors	Timeliness of decision-making	Equilibrium solution
Reference [5]	Single stage	Dynamic game	No	—	Ordinary
Reference [9]	Single stage	Signaling game	No	discrete time	Detailed
Reference [12]	Single stage	Differential game	No	continuous time	Simple
Reference [19]	Multi-stage	Markov game	Yes	discrete time	Ordinary
Reference [20]	Multi-stage	Markov game	Yes	discrete time	Ordinary
This paper	Multi-stage	Markov differential game	Yes	multi-stage continuous time	Detailed

4 Simulation and Analysis

4.1 Description of the simulation environment

The effectiveness and feasibility of the proposed model is verified by simulations using the system shown in Figure 2. The constructed system consists mainly of an external network, a network defense device, an access network, and an internal service network. The network isolation equipment is mainly divided into the firewall, IDS, and PortSentry, and the internal network mainly includes a database server, a file server, and a client terminal.

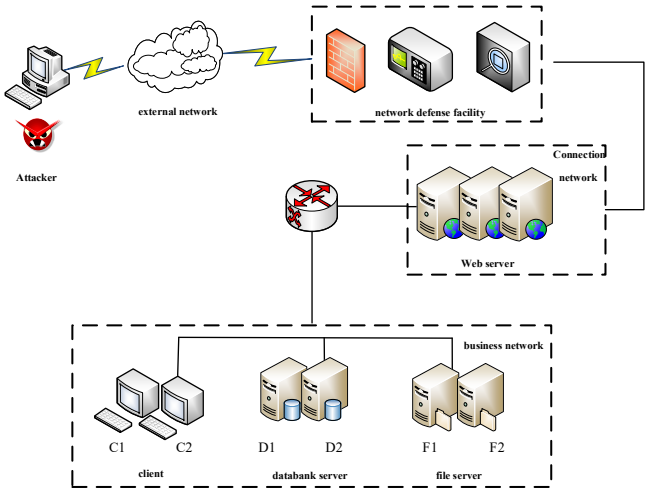


Figure 2: System structure for simulates

The structure of the simulation system is analyzed using the methods in Refs. [21, 22]. The attack and defense process of the moving target is divided into eight stages, as described in Table 2. The specific transfers of the offensive and defensive states are shown in Figure 3. Assuming that the Markov state transfer probability is fixed, we determined the transition probability between different stages using data in the literature and based on expert experience, the specific probabilities are shown in Table 3.

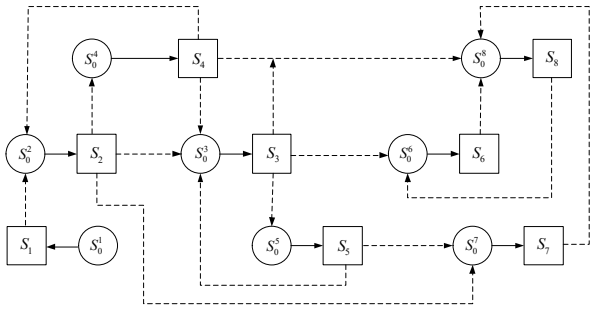


Figure 3: Attack and defense state transfer process

Table 2 States of simulated experimental system at different stages

Attack chain state	State	State description	Attack chain state	State	State description
Scanning detection	S_0^1	Bypass scanning detection tool, acquire root authority	**Implanting attack**	S_0^5	Acquire client C1 access authority
	S_1	Acquire webserver access authority		S_5	Acquire databank server D1 user authority through SQL injection vulnerability
	S_0^2	Acquire webserver root authority through competitive condition vulnerability		S_0^6	Acquire file server F2 access authority
	S_2	Acquire file server F1 access authority		S_6	Acquire databank server D2 user authority through OS command injection vulnerability
Vulnerability exploitation	S_0^3	Acquire file server F1 user authority through cross-site scripting attack XXXS	**Sustaining attack**	S_0^7	Acquire D1 root authority through server replication sub-component security vulnerability
	S_3	Acquire file server F2 access authority		S_7	Acquire sensitive information in databank and cause D1 denial of service
	S_0^4	Acquire databank server D2 user authority through arbitrary document reading vulnerability		S_0^8	Acquire D2 root authority through server replication sub-component security vulnerability
	S_4	Bypass scanning detection tool, acquire root authority		S_8	Trojan planted in databank server D2

Table 3 State transfer probability

State jump	Jump probability	State jump	Jump probability	State jump	Jump probability
$S_1 \to S_0^2$	$\eta(2\mid 1)=0.8$	$S_2 \to S_0^3$	$\eta(3\mid 2)=0.9$	$S_2 \to S_0^4$	$\eta(4\mid 2)=0.8$
$S_2 \to S_0^7$	$\eta(7\mid 2)=0.5$	$S_3 \to S_0^5$	$\eta(5\mid 3)=0.6$	$S_3 \to S_0^6$	$\eta(6\mid 3)=0.9$
$S_3 \to S_0^8$	$\eta(8\mid 3)=0.3$	$S_4 \to S_0^3$	$\eta(3\mid 4)=0.5$	$S_4 \to S_0^2$	$\eta(2\mid 4)=0.6$
$S_4 \to S_0^8$	$\eta(8\mid 4)=0.7$	$S_7 \to S_0^8$	$\eta(8\mid 7)=0.6$	$S_8 \to S_0^6$	$\eta(6\mid 8)=0.8$

According to the access rules of the simulation system, an external network terminal can access the web server, but cannot directly access the file server, the database server, and the client terminal in the internal network. The web server with root authority can access the file server for file sharing. Based on the information security data in the (United States) National Vulnerability Database (NVD), combined with the US National Computer Security Center's offensive and defensive behavior database [23], we constructed an attack strategy set. The strategies are divided into high, medium, and low intensity strategies, as shown in Table 4. The defense strategies are classified according to attribute differences of the system attack surface resources. Table 5 shows the different types of defense strategies in the attack-defense process.

Table 4 Attack strategy at different stages of the game

Stage of game	Attack strategy A	Attack type	Stage of game	Attack strategy A	Attack type
$S_0^1 \to S_1$	Install DLI Trojan	A_H	$S_0^2 \to S_2$	Attack SSH on Web Sever	A_H
	Steal account and crack it	A_M		SQL injection	A_M
	Install socket analyzer program	A_L		SMTP sniffer	A_L
$S_0^3 \to S_3$	Steal account and crack it	A_H	$S_0^4 \to S_4$	SQL injection	A_H
	Send abnormal data to buffer	A_M		Ftp rhost attack	A_M
	Cross-site scripting	A_L		install VBW Trojan	A_L
$S_0^5 \to S_5$	DOM XSS	A_H	$S_0^6 \to S_6$	Oracle TNS Listener	A_H
	CF-exploit attack	A_M		THS chunk overflow	A_M
	Sr-Hard blood	A_L		Ssh buffer overflow	A_L
$S_0^7 \to S_7$	Bash shellshock	A_H	$S_0^8 \to S_8$	Install spy win32 Trojan	A_H
	SSTI	A_M		Shutdown database server	A_M
	Apache chunk overflow	A_L		install SQL Listener	A_L

Table 5 Different types of defense strategy

Defense type	Defense strategy D	Defense type	Defense strategy D
Data type defense strategy	Operands encryption+fixed frequency	Software type defense strategy	ASLR+ fixed frequency
	Fully homomorphic encryption		ISR+ random frequency
	Multiple copy operation program		Multiple variant execution
	Random detection		Code randomization
	Listener detection+ random frequency		Gadget randomization
Network type defense strategy	IP Hopping + random frequency	Platform type defense strategy	SCIT+ random frequency
	POX controller management		TALENT+ fixed frequency
	Port enlarging+ fixed frequency		Multi-configuration migration
	Limit packets		Platform virtualization
	Route changing +random frequency		Platform stack randomization

4.2 Analysis of simulation results

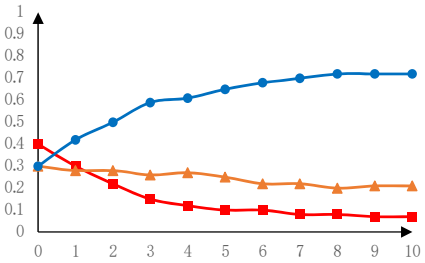
We set the offense-defense return parameter as $r=(8,5,12,7), w=(11,6,10,5)$, the offense-defense strategy cost parameter as $\pi_A=4.2, \pi_D=6.5$, and the discount factor as $\mu(t)=e^{(-0.5(t-t_0^k))}$. The aim of the simulation is to steal and destroy database servers D1 and D2. The algorithm for selecting the strategy selection is implemented using MATLAB2018, and the offense-defense payoffs and the optimal offense-defense strategies $(a_k^*(t), d_k^*(t))$ for each stage are obtained, as shown in

Table 6. The simulation mainly has two attack paths: attack path 1: $S_0^1 \to S_1 \to S_0^2 \to S_2 \to S_0^4 \to S_4 \to S_0^8 \to S_8$ and attack path 2: $S_0^1 \to S_1 \to S_0^2 \to S_2 \to S_0^3 \to S_3 \to S_0^5 \to S_5 \to S_0^7 \to S_7$. Here we take attack path 1 as an example for specific analysis.

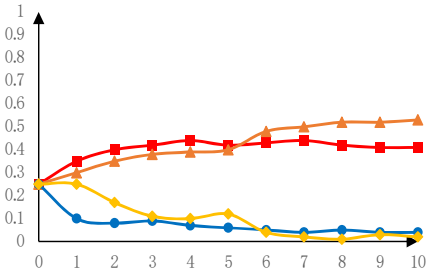
Table 6 Attacker and defender game payoff for different path

$S_0^1 \to S_1 \to S_0^2 \to S_2 \to S_0^4 \to S_4 \to S_0^8 \to S_8$			$S_0^1 \to S_1 \to S_0^2 \to S_2 \to S_0^3 \to S_3 \to S_0^5 \to S_5 \to S_0^7 \to S_7$		
Game stage	Attack payoff	Defense payoff	Game stage	Attack payoff	Defense payoff
$S_0^1 \to S_1$	27.3	15.2	$S_0^1 \to S_1$	31.2	12.7
$S_0^2 \to S_2$	30.2	20.7	$S_0^2 \to S_2$	23.5	21.3
$S_0^4 \to S_4$	45.3	43.9	$S_0^3 \to S_3$	47.2	32.8
$S_0^8 \to S_8$	35.1	34.1	$S_0^5 \to S_5$	63.1	30.7
			$S_0^7 \to S_7$	42.5	25.9

For the first two stages the optimal control trajectory of both offense and defense is shown in Figure 4(a). In the initial stage of the offensive and defensive process, the attack strategy focuses on intelligence gathering and exploratory attacks, and the attacks are mainly of medium and low intensity. The defense strategy mainly focuses on data and network-based defense strategies, which enhances the difficulty for the attacker to collect intelligence. The offensive-defense process belongs to the first and second stages of the attack chain. At the end of this phase the optimal attack strategy is $(0.12, 0.22, 0.6)$ and the optimal defense strategy is $(0.41, 0.53, 0.04, 0.02)$.



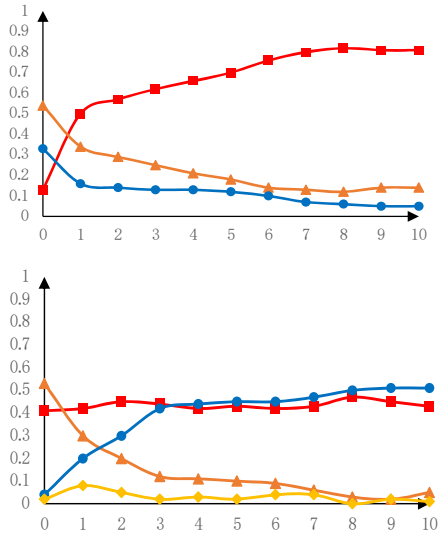
Red: High-level strategy, Brown: Mid-level strategy, Blue: Low-level strategy (Same as below)



Red: data type, Brown: network type, Blue: software type, Yellow: platform type (Same as below)

Figure 4(a): Trajectory of optimal control strategy for first two stages of attack path 1

For $S_0^4 \to S_4$ the attacker began the formal attack based on the intelligence gathering and exploratory attack conducted in the previous stage. The attack strategy is mainly based on high-intensity attacks. The intention is obtain D2 user rights by exploiting the arbitrary file read vulnerabilities and to prepare for the next attack. The defender relies mostly on data and software type defense strategies to prevent further infiltration and damage



by the attacker. This offense-defense process belongs to the third stage of the attack chain, and the optimal control trajectory of both offense and defense is shown in Figure 4(b). At the end of this stage the optimal attack strategy is $(0.8, 0.2, 0)$, and the optimal defense strategy is $(0.43, 0.05, 0.51, 0.01)$.

Figure 4(b): Trajectory of optimal control strategy for third stage of attack path 1

For $S_0^8 \to S_8$ since the goal of the attack is to plant a Trojan horse into the database server D2 to facilitate the next attack, the attack strategy is mainly based on medium-strength attacks. The intention is to obtain D2 root authority by exploiting the server replication sub-component vulnerability and implant a Trojan horse. The defender relies mainly on software- and platform-based defense strategies to prevent the penetration and implantation of the attacker. The offense-defense process belongs to the fourth stage of the attack chain and the optimal control trajectory of both attack and defense is shown in Figure 4(c). At the end of this stage the optimal attack strategy is $(0.1, 0.77, 0.13)$, and the optimal defense strategy is $(0.31, 0.06, 0.01, 0.62)$.

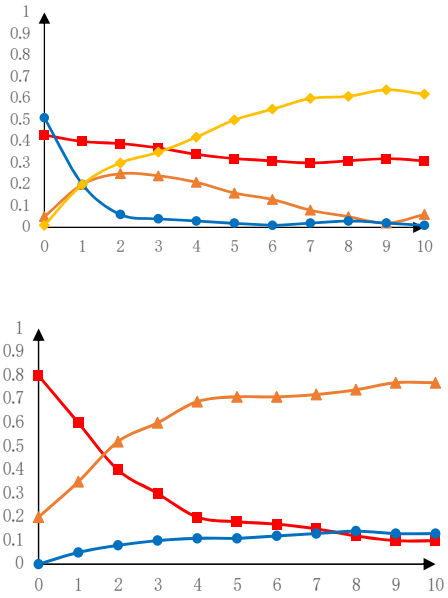


Figure 4(c): Trajectory of optimal control strategy for fourth stage of attack path 1

5 CONCLUSION

As a "game changer" type of new active defense strategy, the moving target defense has enormous potential for development and value for research. In this paper we proposed a model of a dynamic Markov differential game and investigated the method of moving target defense decision-making. When compared to existing methods, the proposed method takes both the real-time nature of defense decision-making and the random nature of the changing security state into account. It provides an effective model for moving target defense research under continuous and real-time conditions. The timeliness of selecting the optimal defense strategy is significant and instructional.

Due to the complexity and diversity of network attack and defense methods, the limited offense-defense strategy set used in this paper cannot fully describe all the variables in an actual offense-defense process. In the meantime the Markov transition probability is mainly based on the limited historical data analysis and expert experience, and there are deficiencies in objectivity and accuracy. These two aspects will be the focus of future research.

ACKNOWLEDGMENTS

We thank all the reviewers for their valuable comments.

REFERENCES

[1] Cybenko G, Wellman M, Liu P, et al. Overview of Control and Game Theory in Adaptive Cyber Defenses[M]//Adversarial and Uncertain Reasoning for Adaptive Cyber Defense. Springer, Cham, 2019: 1-11.

[2] Proactive and Dynamic Network Defense[M]. Springer International Publishing, 2019.

[3] Chen P, Hu Z, Xu J, et al. MTD techniques for memory protection against zero-day attacks[M]//Adversarial and Uncertain Reasoning for Adaptive Cyber Defense. Springer, Cham, 2019: 129-155.

[4] Clark A, Sun K, Bushnell L, et al. A game-theoretic approach to IP address randomization in decoy-based cyber defense[C]//International Conference on Decision and Game Theory for Security. Springer, Cham, 2015: 3-21.

[5] Carter K M, Riordan J F, Okhravi H. A Game Theoretic Approach to Strategy Determination for Dynamic Platform Defenses[C]. ACM Workshop on Moving Target Defense. ACM, 2017.

[6] Prakash A , Wellman M P . Empirical Game-Theoretic Analysis for Moving Target Defense[C]. ACM Workshop on Moving Target Defense. ACM, 2015: 57-65.

[7] Sengupta S , Vadlamudi S G , Kambhampati S , et al. A Game Theoretic Approach to Strategy Generation for Moving Target Defense in Web Applications[C]. International Conference on Autonomous Agents and Multiagent Systems (AAMAS). International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[8] Zhu Q, Başar T. Game-theoretic approach to feedback-driven multi-stage moving target defense[C]. International Conference on Decision and Game Theory for Security. Springer, Cham, 2013: 246-263.

[9] Feng X, Zheng Z, Cansever D. A signaling game model for moving target defense[C]. INFOCOM 2017 - IEEE Conference on Computer Communications, IEEE, 2017: 91-107.

[10] David W. K. Yeung, Leon A. Petrosyan. Differential games theory[M]. New York: Springer Press, 2014.

[11] Nilim A, Ghaoui L E. Active defense strategy selection based on differential game[J]. Operations Research, 2016, 43(12): 163-169.

[12] Hengwei Z, Lv J, Shirui H, et al. Attack-Defense Differential Game Model for Network Defense Strategy Selection[J]. IEEE Access, 2019, 7: 50618-50629.

[13] Changfeng W, Wenying Z. Study on Emergency Decision Making of Major Projects Based on the Dynamic Differential Game Theory[J]. Chinese Management Science, 2017, 25(10): 179-186.

[14] Bing Z. Analysis of Regional Economic Cooperation Behavior Based on Differential Game Theory [J]. Economic Mathematics, 2018 (1): 25-31.

[15] Sengupta S, Chowdhary A, Sabur A, et al. A survey of moving target defenses for network security[J]. IEEE Communications Surveys & Tutorials, 2020.

[16] Zhuang R, Bardas A G, DeLoach S A, et al. A theory of cyber attacks: A step towards analyzing MTD systems[C]//Proceedings of the Second ACM Workshop on Moving Target Defense. 2015: 11-20.

[17] Maleki H, Valizadeh S, Koch W, et al. Markov modeling of moving target defense games[C]//Proceedings of the 2016 ACM Workshop on Moving Target Defense. 2016: 81-92.

[18] Drew F. Jean T. Game Theory [M]. Boston: Massachusettes Institute of Technology Press, 2015.

[19] Doraszelski U, Escobar J F. A theory of regular Markov perfect equilibria in dynamic stochastic games genericity, stability and purification [J]. Theoretical Economics, 2015, 5(2): 369-402.

[20] Lei C, Ma D H, Zhang H Q. Optimal Strategy Selection for Moving Target Defense Based on Markov Game[J]. IEEE Access, 2017, 7: 3274-3286.

[21] Nilim A, Ghaoui L E. Robust control of Markov decision processes with uncertain transition matrices [J]. Operations Research, 2016, 53(5): 780-798.

[22] Doraszelski U, Escobar J F. A theory of regular Markov perfect equilibria in dynamic stochastic games genericity, stability and purification [J]. Theoretical Economics, 2015, 5(2): 369-402.

[23] Gordon L, Loeb M, Lucyshyn W, Richardson R. 2016 CSI/FBI computer crime and security survey[A]. The 2016 Computer Security Institute. San Francisco, USA: IEEE Press, 2016, 48-66.