基于Q学习的DDoS攻防博弈模型研究

史云放 武东英 刘胜利 高 翔
(数学工程与先进计算国家重点实验室 郑州 450002)

摘 要 新形势下的DDoS攻防博弈过程和以往不同,因此利用现有的方法无法有效地评估量化攻防双方的收益以及动态调整博弈策略以实现收益最大化。针对这一问题,设计了一种基于Q学习的DDoS攻防博弈模型,并在此基础上提出了模型算法。首先,通过网络熵评估量化方法计算攻防双方收益;其次,利用矩阵博弈研究单个DDoS攻击阶段的攻防博弈过程;最后,将Q学习引入博弈过程,提出了模型算法,用以根据学习效果动态调整攻防策略从而实现收益最大化。实验结果表明,采用模型算法的防御方能够获得更高的收益,从而证明了算法的可用性和有效性。

关键词 DDoS攻防,矩阵博弈,Q学习,网络熵,纳什均衡
中图法分类号 TP393.08 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.11.040

Research on DDoS Attack-defense Game Model Based on Q-learning

SHI Yun-fang WU Dong-ying LIU Sheng-li GAO Xiang
(State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450002, China)

Abstract The process of DDoS attack-defense game in new situation is different now, so the payoff value cannot be quantified effectively and the game strategy cannot be adjusted dynamically to maximize the payoff using existing methods. In response to this problem, a DDoS attack-defense game model based on Q-learning was designed, and at the same time an algorithm was proposed on the basis of the model. Firstly, the payoff of the attacker and defender was calculated with the network entropy quantitative assessment method. Secondly, the single DDoS attack stage was studied using matrix game method. Finally, the model algorithm was proposed by introducing the Q-learning method into the game process, with which the strategies are adjusted dynamically according to the learning outcomes to maximize the payoff. The result of verification testing shows that the defender can achieve a higher payoff when adopting the model algorithm, thus the algorithm turns out to be practicable and effective.

Keywords DDoS attack-defense, Matrix game, Q-learning, Network entropy, Nash equilibrium

1 引言

随着计算机信息安全技术的发展,网络所面临的威胁层出不穷,网络安全也备受关注。根据Arbor Networks一年一度发布的《Worldwide Infrastructure Security Report》[1],今天的网络运营商面临着非常严峻的安全挑战。报告中指出,分布式拒绝服务(DDoS)攻击日趋复杂,尤其是多矢量DDoS攻击呈上升趋势。因此新形势下DDoS攻击和防御有了新的特点。

目前针对DDoS攻击的检测和防御方法多种多样[2-4]。黄亮等人通过研究神经网络和绩效评估方法,提出了一种防护绩效评估模型,并将其应用至DDoS防护中[5]。以上检测防御方法通常是根据对DDoS攻击流量的建模和特征提取的结果来实施DDoS攻击进行检测和防御的,但其防御策略的选取并没有针对攻击方进行考虑,这使得防御策略难以达到最优。博弈论作为一种决策分析理论,可以从网络攻击和防御两个方面对网络安全问题进行研究,从而越来越受到研究者的关注[6-8]。文献[9]提出一种网络攻防博弈模型研究网络安全测评和最优主动防御。文献[10]针对静态博弈无法满足网络安全防御的现实需求这一问题,提出了适应于完全信息和非完全信息两种场景的动态博弈算法,通过理论分析和实验验证,证明了算法具有很强的可操作性。

通过博弈论分析网络安全问题的关键在于计算攻防的收益。DDoS攻击作为网络攻击的一种,有其自身的特点,所产生的攻击效果不同于其它的网络攻击。本文借鉴网络熵理论的评估量化方法,针对不同攻防策略下的DDoS攻击计算攻防双方所获得的收益,然后运用矩阵博弈对网络攻防进行分析。由于在实际的应用中,攻击初始阶段没有博弈所需要的收益数据,同时只根据一次性的收益评估计算不同攻击阶段的最优策略会产生偏差,有必要根据瞬时收益动态调整策略,因此在矩阵博弈中引入Q学习方法实时调整收益值,减小现实中的不确定性因素引起的误差,从而计算出攻防双方的最优策略。本文首先采用网络熵评估方法对被攻击目标的网络状态进行量化评估,然后提出基于Q学习的博弈模型

到稿日期:2013-11-22 返修日期:2014-02-24 本文受国家自然科学基金(61309007),郑州市科技创新团队项目(10CXTD150)资助。
史云放(1988—),男,硕士生,主要研究方向为信息安全,E-mail:shiyunfang1988@hotmail.com;武东英(1965—),女,硕士,副教授,主要研究方向为信息安全;刘胜利(1973—),男,博士,副教授,主要研究方向为信息安全;高 翔(1984—),男,博士生,主要研究方向为信息安全。

(Q-learning based DDoS Game Model, QDGM)及其算法，最后通过应用实例对模型算法进行实验验证。实验结果表明，采用 QDGM 算法的攻击方或者防御方能够获得更大的攻击收益，证明了算法在实际应用中的可用性和有效性。

2 预备知识

2.1 网络熵

网络熵[11]用于描述网络系统安全性能，可以结合不同的评价指标有效地衡量网络系统的性能变化。根据网络熵理论，熵值越小表明网络系统稳定性越好。系统遭到 DDoS 攻击后，稳定性变差，熵值会增加，所以可以采用熵差对攻击效果进行描述。

对于某个网络熵评估指标，其熵值定义为 $H_i=-\log_2 V_i$，其中 V_i 表示此项网络指标的归一化参数。当网络遭到攻击之后，网络安全整体性能会下降，熵值应该增加，因此采用熵差 $\Delta H=-\log_2(V_2/V_1)$ 对 DDoS 攻击效果进行描述。其中 V_1 表示攻击之前某一指标的归一化性能参数，V_2 为攻击之后的归一化性能参数。

2.2 矩阵博弈

博弈论是研究两个或者两个以上参与者在对抗性局势下，如何采取行动以做出有利于己方的决策的理论。

定义 1 基本的博弈模型可以表示为三元组的形式：$G=\{N_i,S_i,P_i\}$。其中 N_i 表示博弈的第 i 个参与者，S_i 指相应参与者的策略空间，P_i 是指博弈参与者的收益函数，用以计算博弈参与者的收益。

定义 2 满足以下 3 个条件的基本博弈模型称为零和矩阵博弈：

- 博弈参与者的个数为 2，即 $N=\{N_1,N_2\}$。
- 每个博弈参与者的策略空间为有限集。即策略空间 $S_1=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$ 和 $S_2=\{\beta_1,\beta_2,\cdots,\beta_n\}$ 为有限集，并且任从策略空间 S_1 和 S_2 中取出策略 α_i 和 β_j 即可组成策略组合 (α_i,β_j)。
- 对于任意一个策略组合 (α_i,β_j)，博弈参与者的收益分别为 $P_1(\alpha_i,\beta_j)$ 和 $P_2(\alpha_i,\beta_j)$，并且 $P_1(\alpha_i,\beta_j)+P_2(\alpha_i,\beta_j)=0$。

定义 3 零和矩阵博弈的纯策略纳什均衡点 (α_i^*,β_j^*) 是指满足如下条件的策略组合：

$$P_1(\alpha_i,\beta_j^*)\leqslant P_1(\alpha_i^*,\beta_j^*)\leqslant P_1(\alpha_i^*,\beta_j)$$

定义 4 零和矩阵博弈的混合策略纳什均衡是满足如下条件的策略选取概率组合：

$$E(x,y^*)\leqslant E(x^*,y^*)\leqslant E(x^*,y),\forall x\in X_m,\forall y\in Y_n$$

其中，X_m 和 Y_n 表示如下：

$$X_m=\{x=(x_1,x_2,\cdots,x_m)\mid x_i\geqslant 0,i=1,2,\cdots,m,\sum_{i=1}^{m}x_i=1\}$$

$$Y_n=\{y=(y_1,y_2,\cdots,y_n)\mid y_j\geqslant 0,j=1,2,\cdots,n,\sum_{j=1}^{n}y_j=1\}$$

$E(x,y)$ 表示收益期望，$x^*=(x_1,x_2,\cdots,x_m)$ 和 $y^*=(y_1,y_2,\cdots,y_n)$ 表示博弈双方策略选取的概率组合，同时满足 $x^*\in X_m,y^*\in Y_n$。

2.3 Q 学习算法

Q 学习属于强化学习的一种，是一种基于马尔科夫决策过程的在线学习算法，由 Watkins[12] 于 1989 年提出，是目前应用较为广泛的强化学习方法。Q 学习方法使得智能体在不用考虑外部环境模型的情况下，通过对当前的系统状态以及可选的动作做出评价，选择出最优策略，取得收益最大化。

定义 5 马尔科夫决策过程由四元组构成，其定义如下：

$$MDP::=\langle S,A,r,p\rangle \tag{1}$$

其中四元组各部分的含义如表 1 所列。

表 1 马尔科夫决策过程组成部分

组成	含义
S	离散的状态空间
A	离散的行动空间
r	智能体的收益函数
p	智能体的状态转移函数

收益函数 $r:S\times A\rightarrow R$ 表示智能体在某一状态时选择某一动作能够获得的收益。状态转移函数 $p:S\times A\rightarrow[0,1]$ 表示智能体在下一阶段选择某一动作，同时进入到下一个状态的概率。如果智能体在状态转移中获取了瞬时收益 $r(s,a)$，$s\in S,a\in A$，其状态转移概率表示为 $p(s'|s,a)$，则智能体在状态 s 时采取动作 a 所获得的总收益的期望用 Q 值表示如下：

$$Q(s,a)=r(s,a)+\gamma\sum_{s'}p(s'|s,a)Q(s',a) \tag{2}$$

式中，γ 表示折扣因子。智能体根据 Q 学习的结果选择执行动作，而智能体的最优策略就是使得收益最大的策略，表示如下：

$$\pi^*(s)=\max_{s}Q(s,a) \tag{3}$$

智能体在学习过程中不断探索，获取学习经验，调整所执行的动作以取得最大收益。则智能体的 Q 值更新函数可以表示如下：

$$Q_{t+1}(s,a)=(1-\alpha_t)Q_t(s,a)+\alpha_t[r_t(s,a)+\gamma\max_{b}Q_t(s',b)] \tag{4}$$

式中，α_t 为学习速率，在智能体学习过程中为了使算法收敛，α_t 可以随着时间逐渐减小。智能体通过上式不断进行迭代学习，实时更新 Q 值表，调整动作策略。经过一段时间之后，Q 值表趋于稳定，算法收敛，智能体的收益逐步实现最大化。

3 模型设计

3.1 模型思想

通过对 DDoS 攻防博弈过程进行研究，本文建立了基于 Q 学习的 DDoS 攻防博弈模型即 QDGM。在 DDoS 攻防博弈过程中，对攻防双方收益的量化评估尤为重要。由于攻击者所要达到的目的是尽可能使被攻击目标整体性能降低甚至网络瘫痪，而防御者则希望自身网络所受的影响尽量小，因此整个博弈过程可以看作是一种非合作博弈。为了简便，本文将攻防双方的博弈过程当作一种零和矩阵博弈，即攻击方所获取的收益完全来自于防御者的损失。QDGM 模型思想如下：

1)计算攻击者收益。利用网络熵评估方法对攻击前后网络的不同性能指标进行量化评估，在每一个攻击阶段结束后，综合各个评估指标的熵差，定量评估被攻击网络整体的性能下降水平，并将此作为攻击者的收益。

2)衡量被攻击网络的状态。攻击者在实施 DDoS 攻击时，需要实时了解被攻击网络整体的性能情况，即网络的整体状态，针对此状态开展多矢量 DDoS 攻击；对于防御者来说，也需要根据状态来实时调整防御策略。

3)构造特定网络状态的博弈矩阵并计算纳什均衡。选定攻击者和防御者的策略集合，从攻防策略集合中各取一种策

略作为一组攻防策略组合，根据此策略组合下的瞬时收益利用Q学习算法调整矩阵数值，并以此构造出完整的攻防博弈矩阵。根据博弈矩阵，利用线性规划方法计算出纳什均衡。

根据该模型思想，可得QDGM示意图，如图1所示。

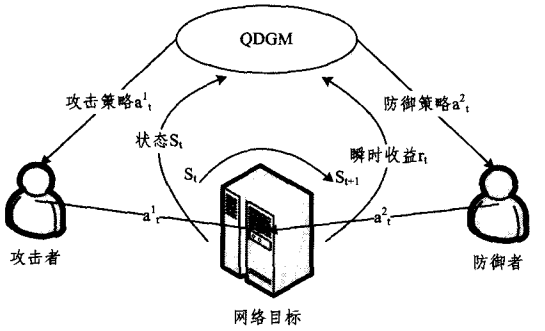

图1 QDGM示意图

3.2 模型描述

根据模型设计思想，提出QDGM的定义。

定义6 基于Q学习的DDoS攻防博弈模型表示如下：

$$QDGM::=\langle S,A^1,A^2,r,p\rangle \tag{5}$$

QDGM总共由5个部分组成。现就模型的各个组成部分进行描述。

• 状态S：模型中的状态是指被攻击网络的整体性能状态，随着攻击过程的进行，状态S也会发生变化。衡量不同网络性能指标可以刻画网络的整体状态。通过网络熵评估方法，引入3种评估指标对网络性能进行评估，它们分别是：网络链路利用率、网络时延和丢包率。下面分别给出此3种指标的评估方法。

1)网络链路利用率是指在某一单位采样时刻网络的数据总量和网络带宽的比值。设攻击前后网络链路利用率分别为U_1和U_2，归一化后结果为V_1和V_2，表示如下：

$$V_1=1-U_1,V_2=1-U_2 \tag{6}$$

利用网络熵对此指标的评估表示如下：

$$\Delta H_V=-\log_2(V_2/V_1) \tag{7}$$

2)网络延迟是指数据包发送和到达的时间间隔，用T表示网络延迟，则归一化的公式如下：

$$E=\begin{cases}\dfrac{T_0-T}{T_0}, & T\leqslant T_0\\ \dfrac{T_0}{T}, & T>T_0\end{cases} \tag{8}$$

式中，E表示归一化的值，T_0是在实际过程中可以接受的最大的延迟时间，分别用T_1和T_2表示攻击前后的网络延迟，则归一化后分别用E_1和E_2来表示，利用网络熵对此指标的评估表示如下：

$$\Delta H_E=-\log_2(E_2/E_1) \tag{9}$$

3)丢包率是指网络访问过程中丢失的数据包占总发送的数据包的比例。在遭到DDoS攻击之后，网络整体性能下降，丢包率会增大。分别用D_1和D_2表示攻击前后的丢包率，L_1和L_2分别表示丢包率的归一化的值，则归一化公式如下：

$$L_1=1-D_1,L_2=1-D_2 \tag{10}$$

则利用网络熵对此指标的评估表示如下：

$$\Delta H_L=-\log_2(L_2/L_1) \tag{11}$$

根据这3种评估指标，可以将网络的整体性能表示出来。为了构造离散的状态空间，本文将以上指标的熵差值按照等级划分进行离散量化，如表2所列。

表2 评估指标离散量化表

ΔH范围	网络性能降低量	量化值
<0.26	<20%	良好
0.26～1.74	20%～70%	中等
>1.74	>70%	较差

将量化值集合用$H_D=\{h_{fine},h_{mid},h_{poor}\}$表示，其中集合中的元素分别表示量化值良好、中等和较差。根据以上3种指标刻画网络整体的性能状态S，可以表示为：$S=\{s=\langle v,e,l\rangle|v,e,l\in H_D\}$。

• 攻击者策略集合A^1：攻击者所具备的DDoS攻击方法的集合。

• 防御者策略集合A^2：防御者为应对DDoS攻击而采取的防御策略的集合。

• 收益函数r：衡量攻防博弈过程中双方的收益。表示形式为：$r:S\times A^1\times A^2\to R$。由上文分析可知，DDoS攻防博弈可以简化为零和矩阵博弈，所以此处仅定义攻击者的收益函数r，防御者的收益函数可以通过攻击者收益函数转换得到。攻击者的收益取决于系统状态、攻击者的攻击策略和防御者的防御策略。收益取值可以用网络熵来衡量，计算公式如下：

$$R=\Delta H=\sum_{i=1}^{3}\omega_i\times\Delta H_i \tag{12}$$

式中，ω表示评价指标对应的权值。ΔH_i表示以上指标在DDoS攻击前后的网络熵差量化值。

• 状态转移函数p：计算在不同的攻防策略下系统发生状态转移的概率，表示形式为：$p:S\times A^1\times A^2\to[0,1]$。随着攻防博弈过程的进行，被攻击网络性能会不断发生变化，状态s发生转移。状态转移函数用以计算在给定的攻防策略下系统状态由s转移到s'的概率。

3.3 基于Q学习的博弈算法

3.3.1 算法描述

对于零和矩阵博弈，Littman等人提出了minmax-Q算法来计算最优策略[13]。在此基础上，本文提出了基于Q学习的博弈算法。基于Q学习的DDoS攻防博弈模型将攻击过程划分为不同的阶段，在攻击阶段开始时，首先利用网络熵评估方法衡量当前被攻击网络的状态，并以此构造本阶段QDGM的状态s。其次，根据状态s以及攻防双方的策略集合A^1和A^2，构造基于Q值的博弈矩阵。再次，根据状态S下的博弈矩阵，利用线性规划算法计算出博弈的纳什均衡点。然后，攻击者或者防御者可以根据纳什均衡点采取相应策略。最后，在下一个攻击阶段到来之时，调整Q值矩阵，准备下一阶段的博弈过程。算法具体如下：

Step 1 初始化阶段计数器$t=0$，利用网络熵评估方法衡量被攻击网络的状态，初始化模型状态为S_0，定义算法的终止条件。

Step 2 对于任意的$s\in S$，初始化其对应的Q值矩阵QM_s。由于任一种DDoS攻击策略都会对网络造成不同程度的影响，因此每一种攻击策略的收益均为非负值。本文假定所有攻击策略对网络造成的性能下降平均为中等程度。查询表2并结合Q值的构造公式初始化矩阵中的Q值。此处根据经验，将值初始化为3，即$QM_s[a_i^1,a_j^2]=3$，其中$a_i^1\in A^1,a_j^2\in A^2$。

Step 3 衡量此阶段的模型状态S_t，根据相对应的Q值矩阵QM_{S_t}，利用线性规划的方法计算此矩阵博弈的纳什均衡

点。此处的计算方法即为模型中的状态转移函数 p。对于攻防双方策略集 A^1 和 A^2，攻防双方的纳什均衡可以用策略的概率分布来表示，分别为：$\Pi^1=(\pi_1^1,\pi_2^1,\pi_3^1,\cdots,\pi_m^1)$ 和 $\Pi^2=(\pi_1^2,\pi_2^2,\pi_3^2,\cdots,\pi_n^2)$，根据此均衡可以实施各自的最优策略。

Step 4　利用网络熵衡量瞬时收益 r_t 以及模型状态，用 s_{t+1} 表示。观察此阶段攻防双方所采用的策略 a_i^1 和 a_j^2。

Step 5　更新状态 s_t 下的 Q 值矩阵 QM_{s_t} 中策略组合 a_i^1 和 a_j^2 所对应的 Q 值，公式如下：

$$QM_{s_t}[s_t,a_i^1,a_j^2]=(1-\alpha_t)QM_{s_t}[s_t,a_i^1,a_j^2]+\alpha_t(r_t+V(s_{t+1})) \tag{13}$$

其中，$V(s_{t+1})=\max\limits_{\pi_i^1\in\Pi^1}\min\limits_{a_j^2\in A^2}\sum\limits_{a_i^1\in A^1}QM_{s_{t+1}}[s_{t+1},a_i^1,a_j^2]\pi_i^1$

Step 6　调整学习速率 α。判断终止条件，如果符合终止条件，算法终止；否则 $t=t+1$，转 Step 3。

3.3.2　*算法分析*

根据以上算法描述，对于每个攻击阶段 t，需要计算 Q 值矩阵的纳什均衡点。设 $|A^1|$ 和 $|A^2|$ 分别表示攻防双方策略集合的大小，则通过线性规划计算纳什均衡点的平均复杂度为 $|A^1|\cdot|A^2|$。通过 Q 学习之后，需要对 Q 值表进行更新，计算更新值的复杂度为 $|A^2|$。假设对于一次完整攻击平均需要 n 个阶段，则算法整体平均复杂度为 $n\cdot|A^1|\cdot|A^2|\cdot|A^2|$。

由此分析可知，算法可以满足攻击者或者防御者在复杂度不高的情况下计算出最优策略的需求。

4　实验以及结果分析

为了对 QDGM 进行验证，本文设计了局域网环境下的 DDoS 攻击实验。实验网络的拓扑结构如图 2 所示。

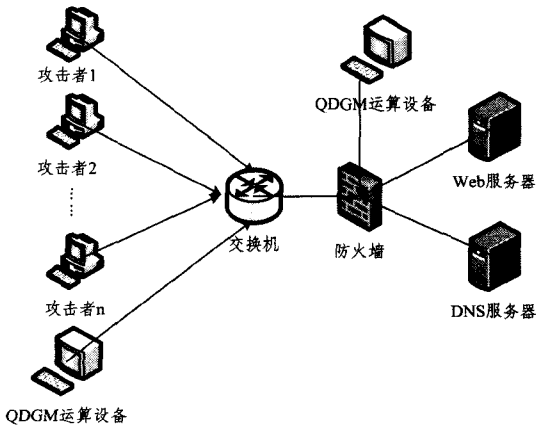

图 2　实验环境下的网络拓扑图

在实际的 DDoS 攻击过程中，为了能够达到攻击效果，攻击源分布在网络的不同位置，并且数目庞大。对于相同的攻击目标，攻击源的数目能在很大程度上影响攻击的效果，但是本文提出的模型只是针对攻防双方的策略博弈而设计的，对攻击源节点数目并无特殊要求，只需要每种攻击策略在无防护条件下均能够达到预期的攻击效果即可。由于是在局域网条件下进行模拟实验，因此被攻击目标的各方面性能都较低，同样对于攻击者而言，攻击者的数目也相对较少。在验证实验中，采用 4 台 HP DL380 服务器共创建了 200 台虚拟机。每台虚拟机配置为：Windows XP 操作系统，256M 内存，1GHz 主频。通过在这些虚拟机中运行特定的攻击程序作为 DDoS 的攻击源。另用 PC 机运行 QDGM 算法程序，同时可以对攻击源发送指令实现对不同攻击策略的转换和调度。采用一台 HP DL388 服务器作为攻击目标，在此服务器上搭建了 Web 服务和 DNS 服务，同时安装了软件防火墙用以配置防御策略。另用一台 PC 机运行 QDGM 算法程序以动态调整防护策略。攻击源机器、QDGM 运算设备和攻击目标逻辑上连接在同一个交换机下，并且在攻击过程中汇聚的总流量总是在交换机能够承载的范围内。

在实验中，为了验证模型的有效性，在防御方采用了模型算法的情况下分别设计了两种 DDoS 打击模式，一种应用了本文中的模型算法来对打击策略进行选取和调整，另一种基于经验来选取打击策略。对于攻击策略，本文选取了已知的几种，如表 3 所列。假定攻击过程中每一阶段只能使用一种攻击策略。

表 3　攻击者策略集合

名称	针对服务	描述
SYN 泛洪攻击	Web	针对目标开放的端口构造大量数据包进行 TCP SYN 连接
TCP session 攻击	Web	针对目标开放端口尝试构造大量全连接
畸形域名攻击	DNS	针对目标 DNS 服务器构造畸形域名进行攻击
UDP 流量攻击	Web DNS	对目标发送大量 UDP 数据包以造成流量拥塞

防御方采用本文中提出的模型算法进行防御策略的选取和调度。防御策略要配置到防火墙的规则中，由 QDGM 运行设备动态调整防火墙规则。实验中选取了已知常用的几种防御策略，如表 4 所列。

表 4　防御者策略集合

名称	描述
SYN Cookie 设置	修改 TCP 3 次握手协议以防范 SYN 泛洪攻击
动态删除连接表项	对当前保持的连接进行有策略的删除
DNS 白名单机制	只提供白名单内的域名解析服务
限制源发包速率	针对特定的源限制其发包速率

防御策略对应的防火墙规则如表 5 所列。

表 5　防火墙配置规则

作用源	作用域	针对协议	规则
TCP 协议修改	Web 服务器	TCP	Allow
TCP 会话表项	Web 服务器	TCP	Delete
DNS 白名单	DNS 服务器	UDP	Allow
源 IP	Web DNS 服务器	TCP UDP	Restrict

下面介绍实验中的两种攻击过程。首先设定状态 s 达到 $s=\langle v=h_{poor},e=h_{poor},l=h_{poor}\rangle$ 或者攻击时间已经达到 30 分钟，攻击过程终止。

对于运用了 QDGM 算法的攻击过程，设每次攻击阶段持续 2 分钟，在攻击开始之前对必要的参数进行初始化。利用网络熵对目标网络进行评估，需要为 3 个评估指标分配权值。此处假定 3 个指标权值均为 1/3，然后通过式(12)计算攻击收益。随着攻防博弈过程的进行，Q 值矩阵会不断更新，我们取其中一个攻击阶段当前状态的 Q 值矩阵，如图 3 所示。

$$\begin{array}{c} \\ a_1^1 \\ a_2^1 \\ a_3^1 \\ a_4^1 \end{array}\begin{array}{c} \begin{array}{cccc} a_1^2 & a_2^2 & a_3^2 & a_4^2 \end{array} \\ \begin{bmatrix} 1.76 & 4.26 & 4.48 & 4.10 \\ 4.22 & 2.70 & 4.32 & 4.32 \\ 4.06 & 4.09 & 1.52 & 4.02 \\ 4.24 & 4.20 & 4.22 & 2.42 \end{bmatrix} \end{array}$$

图 3　阶段 Q 值矩阵

根据纳什均衡理论，任何有限博弈均存在纳什均衡点。由图3所示的博弈矩阵可知，博弈存在纳什均衡点，利用线性规划的方法进行求解，得到混合策略纳什均衡：$\Pi^1=[0.19, 0.32, 0.22, 0.27]$和$\Pi^2=[0.22, 0.36, 0.13, 0.29]$，根据均衡点，攻击者的最优策略是以0.19的概率采用SYN泛洪攻击，以0.32的概率采用TCP session攻击，以0.22的概率采取畸形域名攻击，以0.27的概率采取UDP流量攻击。而防御方的最优策略是以0.22的概率采取SYN Cookie策略，以0.36的概率采取动态删除连接表项策略，以0.13的概率采取DNS白名单机制，以0.29的概率采取限制源发包速率策略。依计算出的概率选择攻击或者防御策略，然后计算瞬时收益，更新Q值矩阵，准备下一阶段的攻击，直至攻击过程结束。

作为对比，本文设计了另一种基于经验选取打击策略的攻击方法。在攻击开始之前，根据以往的攻击经验，TCP Session攻击是针对打击目标开放的TCP端口进行完成3次握手全连接，除了对目标网络进行流量拥塞，最主要的功能是占用TCP的连接数，使正常的访问无法进行TCP连接，同时耗费服务器的资源。所以攻击方选择TCP Session攻击作为打击策略。

对每种攻击过程，利用网络熵方法每隔一定时间对打击目标的网络性能进行评估，绘制了如图4所示的曲线对比图。

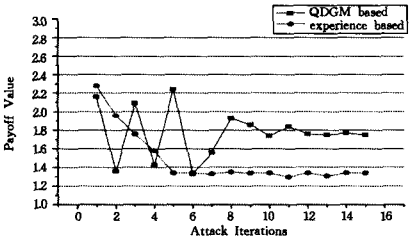

图4 攻击收益对比图

根据收益对比图可知，采用QDGM算法的DDoS攻击由于初始的Q值矩阵和实际的情况差距很大，模型还处于策略选择的“尝试”阶段，导致通过Q学习后Q值矩阵变化较大，所以由此计算出的最优策略会和前一阶段有很大不同。随着攻击的深入，Q值矩阵趋于稳定，最优策略的概率分布也趋于稳定，所以攻击收益也稳定在一定范围内。而在基于经验的DDoS攻击中，由于防御方处于初始阶段，没有应对攻击策略的措施，初始攻击收益较大。随着攻击的持续进行，防御方通过Q学习使得Q值矩阵趋于稳定，防御策略趋于最优化，而攻击策略一直保持不变，使得在稳定阶段攻击收益值低于采用了QDGM算法的攻击方式产生的收益值。

为了对模型的适用性进行验证，本文又针对几种典型的DDoS攻击及其防御设计了如下实验。实验环境网络拓扑图如图2所示，实验方式与上一实验相同。实验中攻击者和防御者的策略集合如表6和表7所列。

表6 攻击者策略集合

名称	针对服务	描述
HTTP泛洪攻击	Web	利用程序模拟代理对目标开放的HTTP服务接进行大量连接
Ping of death攻击	Web DNS	对目标开放端口发送长度大于65535字节的数据包
Teardrop攻击	Web	利用目标处理IP数据包分片重组中的缺陷造成系统资源泄露
Ping泛洪攻击	Web DNS	对目标发送大量Ping数据包以造成流量拥塞

表7 防御者策略集合

名称	描述
添加JS跳转代码	在网页中加入JS代码过滤程序模拟的请求
对数据包长度验证	对每个收到的数据包进行长度以及长度字段检验
IP分片重组预检验	对收到的IP分片在重组前进行预检验从而丢弃异常的IP数据流
限制源发包速率	针对特定的源限制其发包速率

整体的实验过程与上一实验类似，首先是攻防双方均采用QDGM进行实验；然后作为对比，实施对照实验，攻击方基于经验选取打击策略，而防御方采用QDGM算法；最后对比两次实验过程中攻击收益的变化情况并绘制曲线变化图，如图5所示。

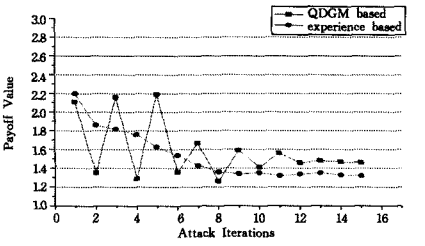

图5 攻击收益对比图

由对比图可知，采取QDGM的攻击方在攻击状态稳定时所获取的收益大于传统的基于经验选取攻击策略的攻击方式，从而进一步验证了模型的有效性。

综合实验结果可知，本文提出的基于Q学习的DDoS攻防博弈模型有较强的适用性。模型算法能够为攻防双方提供策略指导，有一定的实用性。

结束语 本文提出了一种基于Q学习的DDoS攻防博弈模型，模型利用矩阵博弈理论分析DDoS攻防博弈过程，采用网络熵的评估方法对被攻击目标的网络性能进行评估并将评估值作为攻击收益。在实际情况下，由于攻击开始时并没有“经验”供攻防双方参考，并且随着攻击深入，被攻击目标会发生状态迁移，不同的目标状态所适合的攻击和防御策略会有所不同，因此如果要达到最优攻防策略。需要实时衡量收益并以此为反馈来调整策略。本文采用Q学习方法对每一次的博弈过程进行分析，根据瞬时收益和总收益期望计算出最优策略。

针对提出的模型算法，本文设计了实验对算法进行验证。实验结果表明，在防御方使用了模型算法的情况下，采用模型算法攻击方式所获得的攻击收益要大于传统的基于经验的攻击方式的情况，从而证明QDGM算法是有效的。

作为后续，本文将从以下两个方面进行研究：增加网络熵评估指标以便精确刻画被攻击目的状态信息；保存算法运算过程中的数据供将来算法被调用时使用，加快算法的收敛速度从而计算出最优策略。

参 考 文 献

[1] Worldwide Infrastructure Security Report[EB/OL]. [2003-03-01]. http://www.arbornetworks.com/research/infrastructure-security-report

[2] Gupta B B, Misra M, Joshi R C. An ISP level solution to combat DDoS attacks using combined statistical based approach[J]. arXiv preprint arXiv:1203.2400, 2012

（下转第226页）

略描述问题，本文提出了一种 BPEL 流程异常处理策略描述语言 BPEH/PDL，研究了基于着色 Petri 网的 BPEH/PDL 语言的形式化描述方法，并结合“汽车装配流水线管理系统”给出了 BPEH/PDL 语言的实例应用。下一步将继续完善 BPHE/PDL 语言，以提高 BPEH/PDL 的 BPEL 流程异常处理描述能力。

参 考 文 献

[1] Tartanoglu F, Issarny V, Romanovsky A, et al. Coordinated Forward Error Recovery for Composite Web Services[C]// Proceedings of 22nd IEEE Symposium on Reliable Distributed Systems. Florence, Italy, October 2003: 167-176

[2] OASIS. BPEL2.0. Web Services Business Process Execution Language Version 2.0[EB/OL]. http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html

[3] Curbera F, Khalaf R, Leymann F, et al. Exception Handling in the BPEL4WS Language[C]// International Conference on Business Process Management. EINDHOVEN, NETHERLAND, 2003: 26-27

[4] Boutaba R, Aib I. Policy-Based Management: A Historical Perspective[J]. Journal of Network and Systems Management, 2007, 15(4): 447-480

[5] Zeng Liang-zhao, Lei Hui, Benatallah B. Policy-Driven Exception-Management for Composite Web Services[C]// Proceedings of the Seventh IEEE International Conference on E-Commerce Technology (CEC'05). 2005: 355-363

[6] Liu An, Li Qing, Huang Liu-sheng, et al. A Declarative Approach to Enhancing the Reliability of BPEL Processes[C]// 2007 IEEE International Conference on Web Services. 2007: 272-279

[7] Erradi A, Maheshwari P, Tosic V. Recovery Policies for Enhancing Web Services Reliability[C]// IEEE International Conference on Web Services (ICWS'06). 2006: 189-196

[8] Baresi L, Guinea S. A Dynamic and Reactive Approach to the Supervision of BPEL Processes[C]// ISEC'08. 2008: 39-48

[9] Kim K, Choi I, Park C. A Rule-based Approach to Proactive Exception Handling in Business Process[J]. Expert Systems with Applications, 2010, 38(1): 394-409

[10] Chomicki J, Lobo J, Naqvi S. Conflict Resolution Using Logic Programming[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(1): 244-249

[11] Montangero C, Marganiec S R, Semini L. Logic-based conflict detection for distributed Policies [J]. Fundam. Inform. (FUIN), 2008, 89(4): 511-538

[12] Kolovski V, Parsia B, Katz Y. Representing WEB Service Policies in OWL-DL [C]// Proc. of the 4th International Semantic Web Conference (ISWC'05). Galway, Ireland, 2005: 461-475

[13] Kolovski V, Parsia B, Katz Y, et al. Expressing WS policies in OWL [C]// Proc. of WWW 2005 Workshop on Policy Management for the Web. Chiba, Japan, 2005: 29-36

[14] 刘海，刘安，李青，等. 一种 ECA 规则驱动的 BPEL 流程异常处理和分析机制[J]. 小型微型计算机系统，2010，31(7)：1363-1370

[15] Hughes G, Bultan T. Automated Verification of XACML Policies Using a SAT Solver[J]. International Journal on Software Tools for Technology Transfer, 2008, 10(6): 503-520

[16] Huang He-jiao, Kirchner H. Formal Specification and Verification of Modular Security Policy based on Colored Petri Nets [J]. IEEE Transactions on Dependable and Security Computing, 2011, 8(6): 852-865

[17] 孙瑞志，史美林. 工作流异常处理的形式描述[J]. 计算机研究与发展，2003，40(3)：393-397

[18] Hamadi R, Benatallah B. A Petri Net-Based Model for Web Service Composition [C]// Proc. 14th Australasian Database Conf. Database Technologies. ACM Press, 2003: 191-200

[19] Hamadi R, Benatallah B, Mejahed B. Self-adapting recovery nets for policy-driven exception handling in business processes[J]. Distributed and Parallel Databases, 2008, 23(1): 1-44

[20] Peterson J L. Petri Net Theory and the Modeling of Systems [M]. Prentice Hall, 1981

[21] Jensen K. Coloured Petri nets and the invariant method [J]. Theoretical Computer Science, 1981, 14(3): 317-336

[22] Jensen K. Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use[M]. Springer, 1992

[23] Jensen K. Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use[M]. Springer, 1994

[24] Jensen K. Condensed state spaces for symmetrical coloured Petri nets[J]. Formal Methods in System Design, 1997, 9(1/2): 7-40

[25] Jensen K, Kristensen L M. Coloured Petri Nets Modelling and Validation of Concurrent Systems[M]. Springer, July 2009

(上接第 207 页)

[3] Ak M I, George L, Govind K, et al. Threshold Based Kernel Level HTTP Filter (TBHF) for DDoS Mitigation[J]. International Journal of Computer Network and Information Security (IJCNIS), 2012, 4(12): 31-39

[4] 刘陶，何炎祥，熊琦. 一种基于 Q 学习的 LDoS 攻击实时防御机制及其 CPN 实现[J]. 计算机研究与发展，2011，48(3)：432-439

[5] 黄亮，冯登国，连一峰，等. 基于神经网络的 DDoS 防护绩效评估[J]. 计算机研究与发展，2013，50(10)：2100-2108

[6] Bao N, Kreidl O P, Musacchio J. A network security classification game[M]// Game Theory for Networks. Springer Berlin Heidelberg, 2012: 265-280

[7] Bommannavar P, Alpcan T, Bambos N. Security risk management via dynamic games with learning[C]// 2011 IEEE International Conference on Communications (ICC). IEEE, 2011: 1-6

[8] 陈永强，付钰，吴晓平. 基于非零和攻防博弈模型的主动防御策略选取方法[J]. 计算机应用，2013，33(5)：1347-1349，1352

[9] 姜伟，方滨兴，田志宏，等. 基于攻防博弈模型的网络安全测评和最优主动防御[J]. 计算机学报，2009，32(4)：817-827

[10] 林旺群，王慧，刘家红，等. 基于非合作动态博弈的网络安全主动防御技术研究[J]. 计算机研究与发展，2011，48(2)：306-316

[11] 张义荣，鲜明，王国玉. 一种基于网络熵的计算机网络攻击效果定量评估方法[J]. 通信学报，2004，25(11)：158-165

[12] Watkins C J C H, Dayan P. Q-learning[J]. Machine learning, 1992, 8(3/4): 279-292

[13] Littman M L. Markov games as a framework for multi-agent reinforcement learning[C]// ICML. 1994, 94: 157-163