性能优化[J].计算机工程与科学，2020，42（12）：2141-2150.

[9]Clarke I.，Sandberg O.，Wiley B.，et al. Freenet：a distributed anonymous information storage and retrieval system. In：Proc. of the International Workshop on Designing Privacy Enhancing Technologies： Design Issues in Annoymity and Unobservability，Springer-Verlag，New York，USA， 2001.46-66.

[10]Gummadi P.，Saroiu S.，Gribble S.. A measurement study of Napster and Gnutella as examples of peer-to-peer file sharing systems. ACM SIGCOMM Computer Communication Review，2002，32（1）：82-82.

[11]http://en.wikipedia.org/wiki/BitTorrent#Throttling_and_encryption.

[12]Ripeanu M.，Iamnitchi A.，and Foster I.，Mapping the gnutella network，IEEE Internet Computing，6（1），2002.50-57.

[13]Leibowitz N.， Ripeanu M.， and Wierzbicki A.，Deconstructing the kazaa network，Proc.3th IEEE Workshop on Internet Applications. WIAPP 2003. San Jose，CA. USA.，IEEE，2003.112-120.

[14]ARM. Building a Secure System using TrustZone Technology[M]. 2009.

[15]Global Platform Device Technology TEE Client API Specification Version 1.0[EB/OL].2010. http://www.trustedcomputinggroup.org.

[16]Global Platform Device Technology TEE Internal API Specification Version 1.0[EB/OL].2011. http://www.trustedcomputinggroup.org.

[17]Samuel AB，Don F，Virginie G，Franz H，Janne H，Milas F. The trusted execution environment：Delivering enhanced security at a lower cost to the mobile market. White Paper，Global Platform，2011.

[18]Trusted Computing Group. TPM MOBILE with trusted execution environment for comprehensive mobile device security. White Paper，Trusted Computing Group，Incorporated，2012. http://www.trustedcomputinggroup.org.

[19]Atul V. Get into the zone：Building secure systems with ARM TrustZone technology. White Paper，Texas Instruments，2013.

[20]Samsung. An overview of Samsung KNOX. White Paper，Samsung Electronics Co.，Ltd，2013.

[21]Nordström E，Rohner C，Gunningberg P. Haggle：Opportunistic mobile content sharing using search[J]. Computer Communications，2014：121-132.

[22]http://www.peerdevicenet.net/index.html.

[23]Brian Spencer. Making It Easy for Devices to Connect-The All Joyn Peer-to-Peer Architecture [EB/OL].（2011-09-08）.https://developer.qualcomm.com/blog/making-it-easy-devices-connect-%E2%80%93-alljoyn-peer-peer-architecture.

[24]Santos N，Raj H，Saroiu S，Wolman A. Using ARM TrustZone to build a trusted language runtime for mobile applications[J]. ACM Sigplan Notices，2014，49（1）：67-80.

[25]Yang B，Feng DG，Qin Y. A lightweight anonymous mobile shopping scheme based on DAA for trusted mobile platform. IEEE，2014.9-17.

[26]GlobalPlatform Inc. GlobalPlatform Device Technology TEE System Architecture Version 1.1[EB/OL]. White Paper，2014.

[27]A Merlo，L Lorrai，L Verderame. Efficient trusted host-based card emulation on TEE-enabled Android devices[J]. IEEE，2016：454-459.

基于强化学习的网络欺骗防御动态部署研究

◆邵晓　刘曼琳

（海军士官学校　安徽　233012）

摘要：网络欺骗通过在系统内部署虚假的安全弱点，将入侵者引入错误资源达到让其产生错误感知，减少网络安全风险的目的。但传统的网络欺骗防御资源为静态部署，存在着数据收集面较窄、难以适应攻击者的变化等缺陷。本文通过研究基于强化学习的网络欺骗防御动态部署，使用 DQN 算法找寻网络欺骗防御动态部署的最优策略，实现针对网络渗透攻击者的最佳防御效果。

关键词：强化学习；欺骗防御；动态部署

随着信息化和大数据时代的到来，网络空间的博弈也日趋激烈，网络空间安全形势整体比较严峻，网络安全事件频发，网络攻击者的攻击态势也呈现出手段更加先进、方式更加隐蔽、目标更加具体、组织更加严密等诸多变化，配合高技术手段和自动化的攻击装备，导致传统的安全防御手段难以满足现在复杂的网络安全形势。网络欺骗防御通过在网络中部署各种欺骗资源，可以达到干扰网络攻击者的攻击，掩盖网络中的重要目标，识别网络攻击者的攻击手段，延缓攻击者的攻击效果，为最终诱捕提供环境创设条件。人工智能中的强化学习和网络安全防御的结合，为更加有效地防御网络进攻提供了新的方法和途径[1-2]。文献[3]利用强化学习帮助网络防御有效部署 IDS 设备。文献[4]基于强化模型辨别攻击者的攻击路径并以此进行网络漏洞分析。文献[5]将在引入强化学习算法的同时，将攻击图融入攻防博弈模型，设计了一种网络主动防御策略生成方法。本文通过将强化学习模型算法融入网络欺骗防御中，形成网络欺骗防御的智能动态部署中，提高网络防御效果。

1　网络欺骗防御

网络欺骗防御（Deception）是网络防御者通过观察攻击者的网络攻击行为，诱骗攻击者或恶意应用暴露自身攻击意图和攻击手段，以便防御者能据此采取更加有效的防护措施。网络欺骗防御并不等同于传统的蜜罐（Honey Pot）或蜜网技术，除了需要具备与攻击者交互布放诱饵的能力，网络欺骗防御更重要的在于通过伪装和混淆，使用误导、错误响应或其他手段将攻击者诱导至蜜罐中，使其远离重要保护目标，增加攻击者的攻击难度和攻击成本。因此网络防御需要有一个集中管理控制的策略，来创建、分发和管理欺骗资源，如服务器、网络设备、网络应用、网络服务、协议、数据、用户等元素，通过这些元素来诱导吸引攻击者。新的 AI 技术，特别是基于深度学习的强化

学习技术，可以让网络欺骗防御产生与生产环境相匹配的诱饵和欺骗凭证，并能实时自动生成、部署和维持欺骗的进行以及真实性维持，是未来主动防御技术的一个重要的发展方向。

2　强化学习及其模型表示

强化学习是机器学习的一个领域，强调如何基于环境进行行动从而得到最大化的预期利益。通过给定一个马尔科夫决策过程，强化学习寻找一个最优策略，策略就是状态到动作的映射，使得最终的累计回报最大。以单智能体-环境模型为例，单个智能个体和环境之间进行交互，通过操作个体进行决策，来选择相应的操作，操作后环境状态会改变，得到采取动作后的奖励值，如此循环往复，在某个时刻 t 个体采取动作 a 的策略 π 的概率表示为：$\pi(a|s)=P(A_t=a|S_t=s)$，其中 S 表示环境状态，是一个有限的状态集合，S_t 表示在 t 时刻环境的状态，A 表示个体动作，是一个有限的动作集合，A_t 表示 t 时刻个体采取的行动。当出现某个行动 a 后，下一个出现行动的概率为：

$$P_{ss'}^{a}=P(S_{t+1}=s'\mid S_t=s,A_t=a)$$

在策略 π 和状态 s 时，采取价值函数 V_π（s）表示为：

V_π（s）$=E_\pi$（$R_{t+1}+\gamma R_{t+2}+\gamma^2 R_{t+3}+...|S_t=s$）

其中 γ 为衰减因子或折中因子，一般取[0，1]之间，当 γ 取 0 时，表示为贪婪法，当 γ 取 0 至 1 之间的数值则表示当前延时奖励的权重比后续奖励的权重大。

其中 R 为奖励函数表示的是一个期望，环境状态 s 下行动 a 的奖励函数表示为：

$$R_s^a=E(R_{t+1}\mid S_t=s,A_t=a)$$

根据这个模型表示，整个强化学习的马尔科夫决策过程为：

（1）在 t=0 时刻，随机初始化状态 S_0～P（S_0）

（2）智能体根据当前环境状态 S_t 选择行动方案 A_t

（3）智能体采取行动后环境给出奖励 r_t～R（|S_t，A_t）

（4）环境给出下一时刻的状态 S_{t+1}=P（|S_t，A_t）

（5）智能体接收新的奖励 R_t 以及环境状态 S_{t+1}

3　基于强化学习的网络欺骗防御部署策略

3.1 网络欺骗防御部署场景分析

由于网络攻击的状态随时变化，因此网络欺骗防御的部署需要根据当前网络安全状态变化进行智能化调整，在保证网络欺骗防御部署稳定的同时，防止由于攻击者识破欺骗防御资源节点导致部署效能的降低。根据强化学习及其模型表示，将网络防御者看做智能体，当前网络及攻击者看做环境 S_t，实施网络欺骗防御部署动作为 A，防御动作作用于环境后引起网络攻防态势转换将环境转变为 S_{t+1}，同时防御者会得到反馈奖励 R_t，如此反复学习后，最终将得到一个最优策略 π，依据该策略，防御者可以最大可能提高网络欺骗防御效能并对攻击者进行诱捕。在这个过程中，从防御者的角度分析，攻击者的攻击方式主要存在以下两种不确定性，一是网络防御系统的误警或漏警导致对攻击者攻击路径推断的不确定性；二是攻击者对攻击目标的兴趣分布和入侵成功率的不确定性，这两种不确定性也是导致环境状态变化的因素，并且能够在网络欺骗防御网络中进行传导。

3.2 网络欺骗防御部署模型表示

按照强化学习的模型表示以及其决策过程，针对网络环境中的诸多不确定因素，需要通过强化学习来动态部署网络欺骗防御资源，智能选择部署模型，通过强化学习让网络欺骗防御部署策略随着网络环境状态的变化而变化，据此可以将策略表示为：

$$\pi_d^t=\pi_d(S_t)=A_d^t$$

当网络环境为状态 S_t 时，防御者的下一个动作为 A_d^t，通过策略累计的奖赏值 V_π 来判定是否为最优策略，累计奖励可以表示为：

$$V_\pi(S_t)=\sum\gamma^i R_{t+i}$$

可见累计值越大，策略越优即：

$$\pi^*=\text{argmax}V_\pi(S_t)$$

3.3 模型求解

虽然在网络欺骗防御部署场景分析中知道，网络防御系统对入侵检测具有一定的漏警或误警，但考虑到在实际网络攻防环境中遭受的攻击均能为模型的强化学习提供相应的数据，因此通过将多次行动的累计奖励平均值近似为累计奖励的期望值。在网络规模不大，且状态和动作空间离散、维度不高的情况下，可以使用传统的 Q-Learning 算法进行求解。但真实的网络环境复杂多变，为了让策略求解的泛化性能更好，DeepMind 将深度学习技术和强化学习结合，使用 DQN 深度卷积网络 CNN 来逼近值函数，并且利用经验回放训练强化学习的学习过程，通过独立设置了目标网络来单独处理时序差分中的偏差，使得 DQN 在高位连续的状态和动作空间有较好的应用。该算法主要包括以下步骤：

（1）输入状态空间和动作空间，分别用 S 和 A 表示，衰减因子 γ，并设置学习率 α；

（2）初始化经验池 D，容量为 N；

（3）随机初始化 Q 网络参数 Ψ 及 Ψ′；

（4）repeat：初始化网络起始状态 s_0；

（5）repeat：在状态 s_0 选择动作 $a=\pi^\varepsilon$；

（6）执行动作 a，观察网络环境后得到当前动作的奖励 r 以及新的环境该状态 s′；

（7）将 s，a，r，s′放入经验池 D 中进行采样 s_t，a_t，r_t，s_{t+1}；

（8）以 $(y-Q_\Phi(s_t,a_t))^2$ 为损失函数对网络 Q 进行训练，每隔 C 步对网络进行更新；

（9）until：s 为终止状态，返回步骤（5）；

（10）until：∀s，a，$Q_\Phi(s,a)$ 收殓，返回步骤（4）；

（11）输出 Q 网络 $Q_\Phi(s,a)$。

如果只从已知信息中获得最大化奖励而不是从全局的角度出发挖掘环境信息，难以求得最终结果，因此使用 ε−greedy 策略融合探索和利用，以 1-ε 为概率从所有行动中随机抽取 a_t=maxa Q（St，a）。

4　实验

4.1 实验拓扑

为验证策略的有效性，搭建实验拓扑如图 1 所示。攻击者可以通过外部网络进入本网络区域，整个内部网络分为 3 个部分，分别为 DMZ 区域、核心服务区域以及客户终端区域，其中 DMZ 区域主要有一台 Web 服务器 S1 提供 Web 服务，核心服务器有 3 台服务器，分别是文件服务器 S2、数据库服务器 S3、邮件服务器 S4，客户终端区域主要有客户终端 H1 至 H5，其中客户终端 H1 是被网络隔离的。

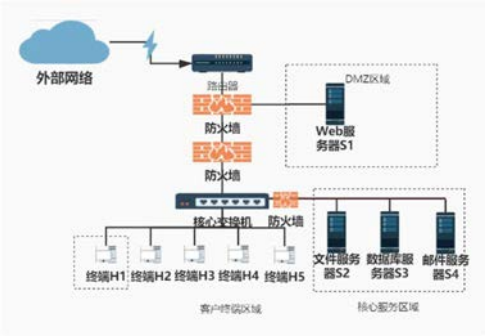

图 1　实验环境拓扑图

4.2 实验方法

根据实验拓扑结构分析，除终端 H1 外，网络中可以有 8 个有效的欺骗资源部署位置，网络初始状态 s_0=[0 0 0 0 0 0 0 0]，网络攻击者的攻击目标为文件服务器 S2，攻击者在内部网络横向移动，网络内部防御中的漏警、误警使用伪随机数生成，模型的训练过程依靠告警数据，与真实网络攻防特点相符，实现对未知网络攻防环境的有效模拟。

在实验中，使用策略 π_d 部署网络欺骗资源，整体防御成功率 ps（π_d）=（num/sum）*100%，其中 num 表示能成功防御网络攻击的次数，sum 为模拟攻防的总次数。

4.3 实验结果及分析

（1）网络欺骗防御静态部署

网络欺骗防御静态部署的实验结果差异性较大，如图 2 所示，不同部署位置、不同的漏警率和误警率下最终的防御成功率 ps 差异较大，最高能达到 71.8%。但在实际网络攻防过程中，静态部署策略结果受到攻击者攻击方式影响较大，难以获得最优的防御策略，导致防御效果很难保持稳定，如果静态部署策略被攻击者识别或侦破，则很难起到网络防御的真正效果。

（2）基于强化学习的网络欺骗防御动态部署

通过强化学习的网络欺骗防御部署，随着强化学习的轮数提升，防御成功率 ps 逐渐提升，最终稳定到 80%以上，且在不同的漏警率和误报率下，其 ps 保持相对的稳定，能够较好地满足不用场合下对网络欺骗防御的部署要求。

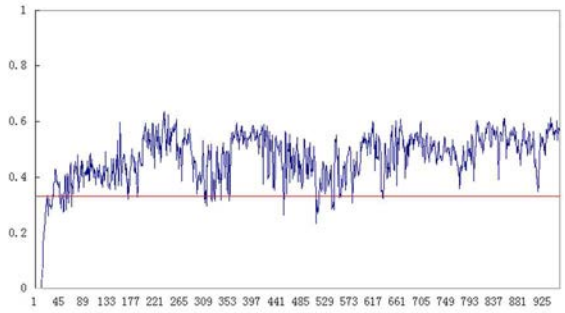

图 2 网络欺骗防御静态部署实验结果

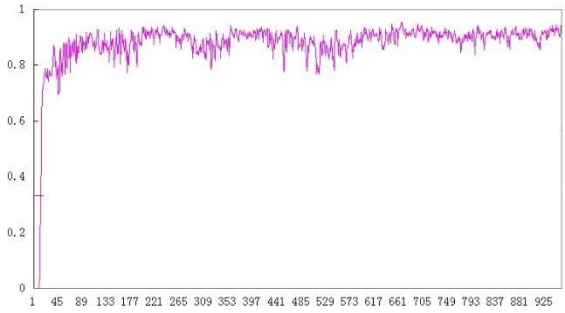

图 3 基于强化学习的网络欺骗防御动态部署实验结果

5 结束语

本文基于强化学习的网络欺骗防御动态部署，根据强化学习模型对网络欺骗防御中的深度学习的场景进行了分析，提出了在强化学习下网络欺骗防御模型算法的实用性，并给出了相应的求解步骤，最后通过仿真实验对策略的有效性进行的验证，理论分析和实验结果均表明，该方法能有效提高网络欺骗防御的成功率，最大可能实现对入侵者的干扰。

参考文献：

[1]王硕.面向多阶段渗透攻击的网络欺骗防御方法研究[D].战略支援部队信息工程大学，2020.

[2]王率.网络欺骗和嗅探技术研究[J].网络安全技术与应用，2013.

[3]Venkatesan S，Albanese M，el.al.Detecting stealthy botnets in a resource-constrained environment using reinforcement learning[C].Proceddings of 4th ACM Workshop on Moving Target Defense.AVM，2017.

[4]Yan J，He H，et al.Q_learning-based vulnerability analysis of smart grid against sequential topology attacks[J].IEEE Transactions on Information Forensics & Security，2017.

[5]金志刚，王新建，李根.融合攻击图和博弈模型的网络防御策略生成方法[J].信息网络安全，2021.

基于分布式服务对网络改造问题的探讨

◆夏邱慧子

（中国民用航空西南地区空中交通管理局通信网络中心　四川　610200）

摘要：笔者于工作中发现，NSX Edge 网关自带的路由功能在报文传输时会发生绕转，防火墙功能也无法满足系统安全性的需求。笔者利用数据层提供的分布式服务改进当前的网络结构，通过部署路由器和分布式防火墙分别替代 NSX Edge 网关的路由和防火墙功能。通过对比分析，分布式服务确实能够优化传输路径、提高系统安全性。

关键词：分布式服务；NSX Edge 网关；网络安全

笔者所在的空管行业中，利用网络虚拟化技术提高系统服务器的资源利用率[1]，通过部署 NSX Edge 网关来管理集群中的虚拟机，并控制不同系统间的访问和流量交互。但 NSX Edge 网关的路由和防火墙功能在现有网络结构中具有局限性，例如：NSX Edge 网关处理容量有限、数据包转发路径绕转、系统的安全性不足等问题。

为满足空管业务飞速发展的需求，笔者认为可以通过部署分布式服务来弥补 NSX Edge 网关的短板。与 NSX Edge 网关的路由功能相比，分布式路由器可以在转发数据包时创建最短的传输路径；而分布式防火墙在扩大处理容量、提高系统安全性方面，相较 NSX Edge 网关自带的防火墙功能具有更好的效果。

1 NSX Edge 网关在生产网络中的应用

相关概念

（1）网络虚拟化

网络虚拟化技术就是在一个物理网络上模拟出多个逻辑网络来满足不同业务的需求，它的逻辑架构主要由四层组成：数据层、控制层、管理层和消费层[2]。

（2）NSX Edge 网关

NSX Edge 网关通常以虚拟机的形式存在，它的作用是连接孤立的网络，并在不同网络间分享上连的网络接口，可以实现二层桥接、三层路由、防火墙等功能[3]。

（3）分布式路由器

分布式路由器通常以虚拟机的形式存在，作为分布式路由的控制单元，主要处理从虚拟机到虚拟机的三层流量。

（4）分布式防火墙

分布式防火墙通常扮演边界防火墙的角色，阻止未授权的用户访问受保护的网络，主要用来处理东西向流量。与传统防火墙相比，它还可以扩充东西向防火墙的容量，提供更加精细颗粒度的访问控制。

1.1 NSX Edge 网关在生产业务中的应用

分布式服务和 NSX Edge 网关服务主要运行于数据层，通常分布式服务主要用于控制虚拟机之间东西向流量的交互，而 NSX Edge 网关服务作为物理网络和虚拟网络的接口，用于处理南北向流量的交互。

在实际生产业务中，利用 NSX Edge 网关的路由功能和防火墙功能管理各系统集群下的虚拟机及各系统之间的流量交互。以双流机场为例，图 1 所示为双流机场系统网络连接拓扑图。A 系统服务器和 B 系统服务器各自管理服务器上的虚拟机，而系统服务器上连服务器汇聚交换机和核心交换机，并通过服务器汇聚交换机互联、核心交换机互联实现冗余备份，数据包通过核心交换机到达系统外部防火墙，经外部防火墙策略过滤后最终与外部用户实现交互。

对于 A 系统和 B 系统之间的流量交互，主要是通过 NSX Edge 网关来实现的，A 系统虚拟机将数据发往 A 系统 NSX Edge 网关，经过 NSX Edge 网关的路由选择和防火墙策略控制到达 B 系统 NSX Edge 网关，最后发往 B 系统虚拟机。从服务器到服务器的访问，都是数据中心内部的东西向流量，这与部署在数据中心边界的防火墙设备无关，需要通过 NSX Edge 网关配置防火墙策略来实现控制[4]。

在实际运用中，为了实现弹性和高可用性，通过网络虚拟化平台将 NSX Edge 网关配置为双台，即 Active/Standby（A/S）冗余部署模式，如图 1 中红色虚线方框内所示。当其中一台 NSX Edge 网关处于 Active 状态时，作为主用 Edge 网关承载网络中的流量；而另一台 NSX Edge 网关则处于 Standby 状态，作为备用 Edge 网关，不承载网络中的流量和服务，但在主用 Edge 网关失效后接管其工作。对于信息同