文章编号:1671-6833(2026)04-0108-09

突变-服务欺骗协同的移动目标防御方法

张建辉[1,2], 徐思捷[1], 曾俊杰[1], 王瑞民[3]

(1.郑州大学 网络空间安全学院,河南 郑州 450002;2.嵩山实验室,河南 郑州 450046;3.郑州大学 计算机与人工智能学院,河南 郑州 450001)

摘 要:针对数字孪生网络(DTN)中突变类移动目标防御(MTD)策略因离散触发而难以在触发间隔内持续拦截恶意流量,易形成防御空窗的问题,提出一种突变-服务欺骗协同的 MTD 方法(MSD-MTD)。在地址突变和服务端口突变基础上,引入服务欺骗机制对突变间隔内的可疑流量进行重定向,以增强持续防护能力;进一步结合基于跨节点流量对齐与特征选择的入侵检测方法感知网络状态,并利用深度 Q 网络(DQN)实现 MTD 策略的自适应选择。在 Mininet-WiFi 平台上,基于 CICIDS-2017、CICIDS-2018 和 UNSW-NB15 数据集开展对比实验,并与两种典型地址突变方法进行比较。结果表明:MSD-MTD 在 3 个数据集上的平均防御成功率分别达到 93.36%、88.20% 和 95.50%,且往返时延主要分布在 0~2 ms,说明所提方法在提升防御效果的同时对网络服务时延影响较小。

关键词:数字孪生网络; 移动目标防御; 服务欺骗; 深度强化学习

中图分类号:TP302.1;TP302.7 **文献标志码**:A **doi**:10.13705/j.issn.1671-6833.2026.04.019

数字孪生网络(digital twin network,DTN)是物理网络的数字映射体,可通过数据驱动模型实现状态预测与反馈优化[1],并已在未来网络、智慧城市和交通出行等场景得到应用[2]。但虚实映射与双向交互也扩大了系统攻击面[3]。Wang 等[4]将 DTN 划分为感知层、网络层、传输层、运行层和应用层。在该架构下,感知层作为 DTN 与物理网络的交界面,易受中间人攻击、拒绝服务攻击(DoS)等威胁。已有研究尝试引入深度学习识别异常数据,但仅依赖检测仍难以保障 DTN 安全:一是检测存在时滞以及误报、漏报;二是攻击流量可通过低慢小特征或对抗样本规避检测,因此仍需主动防御机制在早期削减攻击面并切断攻击链[5]。

移动目标防御(MTD)通过持续扰动系统状态使攻击者侦察结果快速失效,是一种典型的主动防御技术[6]。按触发方式,MTD 可分为定时触发与事件触发两类;在策略层面,Cho 等[7]将其归纳为随机化、多样化和冗余 3 类,既可独立应用,也可组合形成混合式 MTD。

针对触发方式的研究,近年来相关工作多以事件触发为主,并常与时序预测模型结合,以提高防御动作与安全事件之间的匹配程度。Zhang 等[8]面向数字孪生移动网络提出协同突变型 MTD 方法,首先利用长短期记忆网络(LSTM)预测未来安全事件,并将预测结果映射为状态表征;随后将突变策略的联合部署表示为半马尔可夫决策过程,在兼顾防御收益与调频开销的前提下,提高了防御效果与执行效率。

在策略层面,现有研究多从防护收益与资源开销两个方面评估 MTD 策略。Rehman 等[9]面向物联网(IoT)提出融合操作系统多样化与网络欺骗的主动防御框架,并进一步设计了基于重要性度量的操作系统多样化方法,以降低防御开销。Masud 等[10]基于 3 层时间分层攻击表示模型构建 IoT 混合 MTD 框架,采用随机化、多样性与冗余等防御策略,对不同策略组合下的攻击成功概率、攻击风险和攻击成本等指标进行了量化分析。Zhou 等[11]面向边缘云环境提出协同 MTD 机制,并利用深度强化学习进一步优化资源利用率。扈红超等[12]针对云原生环境中的正则表达式拒绝服务(ReDoS)攻击,提出了基于 MTD 的防护方法,并在 Kubernetes 环境下通过关

收稿日期:2026-03-29;**修订日期**:2026-05-10

基金项目:国家重点研发计划(2023YFB2906401);嵩山实验室资助项目(221100210900)

作者简介:张建辉(1977—),男,河南平顶山人,郑州大学副研究员,博士,主要从事网络空间内生安全、新型网络架构和网络数字孪生等研究,E-mail:ndsczjh@163.com。

引用本文:张建辉,徐思捷,曾俊杰,等. 突变-服务欺骗协同的移动目标防御方法[J]. 郑州大学学报(工学版),2026,47(4):108-116. [Zhang Jianhui, Xu Sijie, Zeng Junjie, et al. A mutation-service deception collaborative moving target defense method[J]. Journal of Zhengzhou University (Engineering Science),2026,47(4):108-116.]

键微服务识别与动态轮换提升了防护的主动性和效率。Tan 等[13]综述了博弈论在 MTD 中的应用,指出在复杂攻防交互条件下,如何实现科学而有效的策略选择是 MTD 研究中的关键问题。随着相关研究的推进,欺骗策略因构建成本较低、攻击面扩展灵活,逐渐成为 MTD 研究的重要方向之一。

尽管相关研究已取得进展,但仍存在两点不足:一是多数工作更关注移动方式,而对触发时机的设计关注不足;二是 DTN 攻击场景仍以突变和随机化策略为主。以智能驱动的主机地址突变[14](IDHAM)为例,通过马尔可夫决策过程(MDP)实现主机地址突变并保持会话连续,但突变策略仍难以避免触发间隔内的防御空窗,攻击者可能沿用上一轮侦察情报继续发起攻击。

为解决上述问题,本文提出突变-服务欺骗协同的移动目标防御方法(MSD-MTD),在事件触发框架下将突变策略的阶段性扰动与服务欺骗的持续承接相结合:当突变策略在触发间隔内仍难以持续阻断可疑访问时,利用服务欺骗将可疑流量重定向至轻量级欺骗节点,以填补防御空窗;同时引入网络状态感知与自适应决策机制,在防御收益与网络服务时延影响之间进行动态权衡。本文的创新点主要体现在以下两个方面。

(1)MSD-MTD 方法。在常规 DTN 中,突变策略主要改变攻击暴露面,而本文进一步引入服务欺骗机制,在突变尚未再次触发时,将可疑流量重定向至欺骗节点,对间隙期攻击流量进行分流,从而减小防御空窗。

(2)事件驱动的 MDP 建模与策略选择的实现。针对 DTN 多源流量在时间和空间上的分散性,先进行跨节点时间戳对齐与特征筛选,再利用 ModernTCN 识别安全事件,并以此作为网络状态;在此基础上,将防御部署过程表示为事件驱动的 MDP,并采用深度 Q 网络(DQN)在防御收益与网络服务时延影响之间进行权衡,实现策略的自适应选择。

1　威胁模型

图 1 为基于 SDN 的 DTN 攻击场景示例,在基于软件定义网络(SDN)的 DTN 中,攻击者可经由接入交换设备向目标服务节点发起攻击。本文假设 DTN 的核心基础设施与关键服务平台均可信且未被攻陷,潜在攻击者位于外部网络,无法直接获取内部网络状态或篡改控制平面。

攻击者通常先对目标进行侦察,获取漏洞节点、开放端口等信息,随后发起攻击。以 DoS 为例,攻击者通过操控僵尸网络向目标发送大规模恶意流量,造成节点资源耗尽或服务中断。攻击结束后,攻击者可依据反馈结果再次进入侦察阶段,并为下一轮攻击调整目标参数。

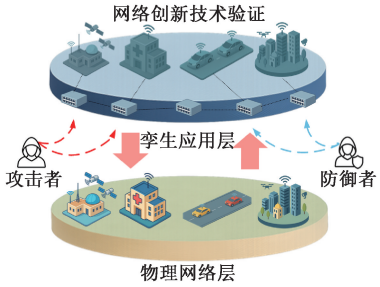


图 1　基于 SDN 的 DTN 攻击场景示例

Figure 1　An example of DTN attack scenario based on SDN

假定对手具备一定适应能力但资源受限,僵尸网络难以长期维持高强度持续攻击,攻击行为呈间歇式。在攻击间歇期内,攻击者可根据新一轮侦察结果动态调整目标 IP、端口及连接参数,以尝试规避目标系统中新部署的 MTD 策略。

2　MSD-MTD 方法设计

现有工作多采用地址突变、路由突变和端口突变等策略。此类方法虽能扰乱攻击链、延缓攻击推进,但仍受触发间隔和资源开销限制,既难以在触发间隙持续拦截恶意流量,也可能因高频触发带来额外重配置代价并引发控制面震荡。因此,单纯依赖突变策略难以满足 DTN 的综合防御需求。为弥补其持续防护不足,本文面向基于 SDN 的 DTN 攻击场景提出 MSD-MTD 方法。

2.1　MSD-MTD 总体架构

图 2 给出了 MSD-MTD 方法的总体架构。

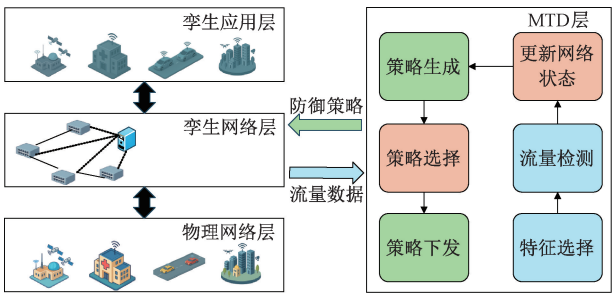


图 2　MSD-MTD 总体架构

Figure 2　MSD-MTD overall architecture

系统由物理网络层、孪生网络层、孪生应用层和 MTD 层 4 部分构成。各 DTN 服务节点持续采集流量数据并上报至 MTD 层。MTD 层首先对汇聚日志进行预处理,并在统一时间轴上完成对齐与特征筛

选。随后,入侵检测模块对安全事件进行识别,并将结果映射为 DTN 网络状态,为后续防御决策提供依据。MTD 层包含策略选择模块和策略下发模块。策略选择模块基于 DQN 输出联合动作,策略下发模块负责将防御动作解析为可执行的网络操作,并由 SDN 控制器通过南向接口向交换设备下发转发表项和重定向规则。上述流程在攻击周期内循环执行。

2.2 突变-服务欺骗协同策略

在 DTN 中,突变策略虽可通过改变对外配置扰乱攻击节奏,但在相邻两次配置切换之间仍存在防御空窗,攻击者可沿用既有侦察情报继续发起攻击。为弥补这一空窗,本文在突变策略基础上引入服务欺骗机制,通过控制面下发高优先级规则,将指向失效配置的可疑流量重定向至隔离的欺骗节点,从而削弱攻击效果。

网络欺骗技术通常包括蜜罐、混淆、诱饵和重定向等形式[15]。相比将 MTD 与多种欺骗机制联合部署的方法,本文所提方法采用轻量级服务欺骗节点承接攻击流量,以降低资源消耗并便于与突变策略协同。

本文将节点负载能力定义为在给定业务服务与网络条件下,节点在一个时间窗内可稳定处理的最大请求数 C。当输入负载持续上升并导致往返时延(RTT)超过阈值时,对应负载视为容量上限。设真实服务节点容量为 C_r,欺骗节点容量为 C_d,并取 $C_d = 0.5C_r$,以体现欺骗节点的轻量化部署特性。

欺骗节点的启用数量由 DQN 策略自适应确定。系统在每个时间窗结束后根据当前防御效果与网络服务时延影响计算奖励,并据此调整下一时刻启用的欺骗节点数量;随后由 SDN 控制器同步更新链路状态与转发表规则,使可疑流量按策略重定向至相应欺骗节点。

如图 3 所示,在突变策略的防御空窗期间,系统将指向失效配置的可疑流量重定向至轻量级欺骗节点。正常用户按最新配置接入,服务访问不受影响。

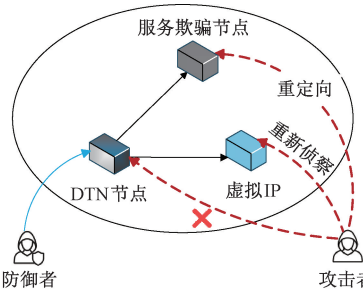


图 3 突变-服务欺骗协同机制示意图

Figure 3 Schematic diagram of the mutation-service deception collaborative mechanism

以一次突变策略触发到下一次触发为 1 个周期,策略下发流程如图 4 所示。

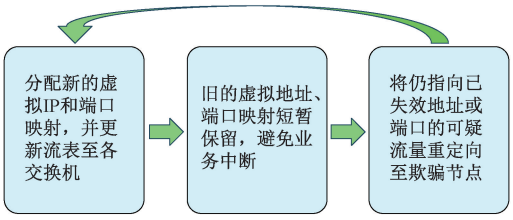


图 4 MSD-MTD 策略执行与规则下发流程

Figure 4 MSD-MTD strategy execution and rule-deployment workflow

控制器首先为指定节点分配新的虚拟 IP 与服务端口映射并下发流表;在短暂过渡阶段,旧的虚拟地址、端口映射暂时保留以避免业务中断;当旧配置失效后,仍指向已失效地址或端口的流量被视为可疑流量,并统一重定向至欺骗节点。与传统蜜罐相比,该机制无须暴露独立诱饵服务,且欺骗节点部署在隔离分区,仅处理被引流的可疑连接,因此资源开销更低、对网络影响更小。

随着策略执行过程中流表项不断增加,交换机查询效率会下降,因此需对其可用流表容量加以约束:

$$B_j + \sum_{f \in F_t} \delta_{f,j} \leqslant C_{j,\max}, \quad \forall j \in \mathcal{S}。 \tag{1}$$

式中:$\mathcal{S}$为 OpenFlow 交换机集合;j 为其中任一交换机;$C_{j,\max}$ 为交换机 j 的流表最大可用容量;B_j 为常驻规则已占用的流表项数量;F_t 为周期 t 内需要添加流表的流量集合;$\delta_{f,j}$ 为指示量,若流量 f 的路径经过交换机 j 且需要在其上安装规则,则为 1,否则为 0。

3 MSD-MTD 建模

基于设计的 DTN 攻击场景,将安全事件与 MTD 策略部署建模为事件驱动的 MDP。与传统 MDP 不同,事件驱动的 MDP 的状态转移由入侵检测输出直接驱动,包含状态空间、动作空间、状态转移以及奖励函数,具体设计如下。

3.1 状态空间

设 DTN 中节点编号集合为 $\mathcal{N} = \{1, \cdots, n\}$,其中 n 表示网络中最大节点数量;安全事件集合为 $E = \{e_0, \cdots, e_{k-1}\}$,其中 e_0 表示良性流量,其余为不同类型的恶意流量,共有 k 种安全事件。对于任一时间窗 t,入侵检测模块 ModernTCN 对节点 i 输出的安全事件记为 $e_{i,t} \in E$。据此,本文将时间窗 t 的网络状态定义为由所有节点安全事件共同组成的状态 $S_t = (e_{1,t}, \cdots, e_{n,t})$。由于每个节点在每个时间窗仅对应一个离散事件标签,且安全事件类别数为 k,因此 n

个节点组成的网络状态总数为 $\mu = k^n$,由此可以将 MDP 的状态空间描述为 $\{S_1, \cdots, S_\mu\}$。

为保证后续基于 DQN 的策略网络能够直接处理该状态信息,本文进一步将 S_t 从节点安全事件集合编码为定长数值。具体而言,首先为安全事件集合 E 中每一类事件分配唯一编号 $\mathrm{id}(e)$。随后,对节点 i 在时间窗 t 的事件 $e_{i,t}$ 构造 one-hot 向量 $\boldsymbol{o}_{i,t}$,其分量定义为

$$o_{i,t}^{(r)} = \begin{cases} 1, & r = \mathrm{id}(e_{i,t}); \\ 0, & r \neq \mathrm{id}(e_{i,t}). \end{cases} \tag{2}$$

式中:$r = 0, 1, \cdots, k-1$。然后将所有节点的 one-hot 向量按节点编号顺序拼接,得到 DQN 的状态输入向量:

$$\boldsymbol{s}_t = [\boldsymbol{o}_{1,t}^{\mathrm{T}}, \boldsymbol{o}_{2,t}^{\mathrm{T}}, \cdots, \boldsymbol{o}_{n,t}^{\mathrm{T}}]^{\mathrm{T}}。 \tag{3}$$

3.2　动作空间

在时间窗 t 内,防御者可对不同节点分配不同的 MTD 策略,包括 IP 地址突变、服务端口突变和服务欺骗 3 种。策略集合记为 $\Omega = \{p, s, d\}$,其中 p 表示 IP 地址突变,s 表示服务端口突变,d 表示服务欺骗。根据网络状态 S_t 的定义,可将联合动作表示为 $A_t = \{\boldsymbol{a}_{1,t}, \boldsymbol{a}_{2,t}, \cdots, \boldsymbol{a}_{n,t}\}$。节点 i 在时间窗 t 内执行的动作表示为三维向量 $\boldsymbol{a}_{i,t} = \left(a_{i,t}^p, a_{i,t}^s, a_{i,t}^d\right)$,其中 $a_{i,t}^p$ 与 $a_{i,t}^s$ 为二值决策,分别表示是否触发 IP 地址突变与服务端口突变。服务欺骗策略对应的分量不限定为二值,而用离散等级表示欺骗节点规模,记为 $a_{i,t}^d \in \{0, 1, \cdots, D_{\max}\}$,其中 0 表示不启用服务欺骗,数值越大表示更多的欺骗节点,$D_{\max}$ 为节点最大数量。具体动作由深度强化学习算法在综合防御收益与网络服务时延影响的权衡下确定,例如 $(1,1,0)$ 表示节点 i 在时间窗 t 内同时执行 IP 地址突变与服务端口突变但不启用服务欺骗,$(0,0,0)$ 表示不执行任何动作。

3.3　状态转移

状态转移由入侵检测输出的安全事件序列驱动。系统在每个时间窗结束后,基于该时间窗内各节点的流量日志生成时间窗样本,并由 ModernTCN 输出各节点的安全事件类别。随后,将每个节点在该时间窗内的检测结果进行汇总,得到该节点在当前时间窗对应的安全事件标签,并按节点编号顺序组合形成当前时间窗的网络状态 $S_t = (e_{1,t}, \cdots, e_{n,t})$。

3.4　奖励函数

在 MDP 中,奖励函数用于度量智能体在时刻 t 采取动作后在下一决策周期产生的即时收益与代价,从而为策略优化提供目标信号。本文的奖励由两部分构成:防御奖励 R_{d} 与网络服务时延影响惩罚项 R_{c}。

防御奖励定义为

$$R_{\mathrm{d}} = -\alpha \sum_{i \in N} \lambda_{t,i} + \beta \sum_{i \in N} v_{t,i}。 \tag{4}$$

式中:$\alpha = 1$;$\beta = 1$;$\lambda_{t,i}$ 为节点 i 上不期望结果对应的流量总量,即恶意流量成功到达与良性流量被误阻断的总和;$v_{t,i}$ 为节点 i 上期望结果对应的流量总量,即恶意流量被成功阻断与良性流量成功到达的总和。

为抑制过渡策略触发带来的网络服务时延影响,引入网络服务时延影响惩罚项:

$$R_{\mathrm{c}} = -\eta \sum_{i \in N} \kappa_{t,i}。 \tag{5}$$

式中:$\eta = 0.3$;$\kappa_{t,i}$ 为节点 i 触发 MTD 策略带来的 RTT 增量。

综上,完整的奖励函数为

$$R_{\mathrm{t}} = R_{\mathrm{d}} + R_{\mathrm{c}}。 \tag{6}$$

4　入侵检测与策略选择算法

4.1　基于跨节点流量对齐与特征选择的入侵检测

基于跨节点流量对齐与特征选择的入侵检测流程如图 5 所示。

首先,对各节点产生的原始流量日志进行统一预处理。考虑到不同节点日志起始时间可能存在偏

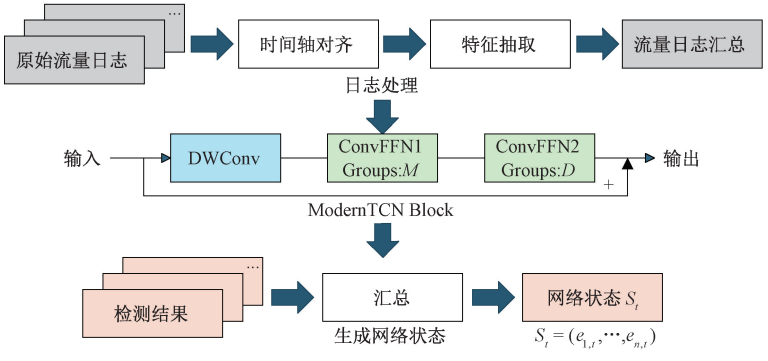


图 5　基于跨节点流量对齐与特征选择的入侵检测流程

Figure 5　Intrusion detection process based on cross-node traffic alignment and feature selection

移,本文以控制器时间为参考对节点日志时间戳进行校准,并将各节点流量记录映射到统一时间轴。最后,按固定步长划分时间窗,对落入同一时间窗的流量记录进行聚合统计,构造定长特征向量,用于表示该节点在该时间窗内的通信状态。

为降低冗余特征带来的噪声并提升模型判别效率,本文采用 scikit-learn 中的 SelectKBest 方法[16]进行特征筛选,在训练集上依据特征与标签的相关性评分筛选保留前 10 个特征。特征选择与标准化流程如算法 1 所示,选择器与标准化器均仅在训练集上拟合,并一致应用于训练和测试集,以避免信息泄漏并保证特征量纲一致。

算法 1 特征选择与标准化流程。

输入:训练样本特征矩阵 $\boldsymbol{X}_{\mathrm{tr}}$、标签 $\boldsymbol{y}_{\mathrm{tr}}$、测试特征矩阵 $\boldsymbol{X}_{\mathrm{te}}$、保留特征数 K、评分函数 f_{classif};

输出:变换后的训练特征矩阵 $\boldsymbol{X}_{\mathrm{tr}}$、测试特征矩阵 $\boldsymbol{X}_{\mathrm{te}}$、特征选择器 FS、标准化器 SC。

① FS←SelectKBest(score_func=f_{classif},$k=K$);

② FS.fit($\boldsymbol{X}_{\mathrm{tr}}$, $\boldsymbol{y}_{\mathrm{tr}}$);

③ $\boldsymbol{X}_{\mathrm{tr}}$=FS.transform($\boldsymbol{X}_{\mathrm{tr}}$);

④ $\boldsymbol{X}_{\mathrm{te}}$=FS.transform($\boldsymbol{X}_{\mathrm{te}}$);

⑤ SC=StandardScaler();

⑥ SC.fit($\boldsymbol{X}_{\mathrm{tr}}$);

⑦ $\boldsymbol{X}_{\mathrm{tr}}$=SC.transform($\boldsymbol{X}_{\mathrm{tr}}$);

⑧ $\boldsymbol{X}_{\mathrm{te}}$=SC.transform($\boldsymbol{X}_{\mathrm{te}}$)。

本文采用 ModernTCN[17]作为入侵检测模型的骨干网络,其结构如图 5 所示。ModernTCN 的输入为按时间窗组织后的多变量时间序列特征,模型通过深度卷积与卷积前馈结构分别建模时间依赖、变量内特征重组以及跨变量信息交互,从而获得对网络攻击事件更具判别力的表示。具体而言,ModernTCN Block 由 3 部分协同完成信息提取与融合:DWConv 仅沿时间维进行卷积建模,各通道之间互不交叉,用于捕获时间相关性并扩大有效感受野;ConvFFN1 用于变量内的特征重组与增强;ConvFFN2 用于跨变量的信息交互与融合。

随后,对各节点在该时间窗内产生的多个样本判别结果进行统计:若某节点在该时间窗内无恶意流量记录,则沿用上一时间窗状态;若存在恶意流量,则以该时间窗内出现频次最高的攻击类别作为该节点的最终安全事件。最后,将该时间窗内各节点的分类结果按编号顺序汇总,得到 DTN 的网络状态 $S_t=(e_{1,t},\cdots,e_{n,t})$。

4.2 基于 DQN 的 MTD 策略自适应选择

为实现 MSD-MTD 策略的自适应选择,本文采用深度 Q 网络[18](DQN)。传统的 DQN 将智能体与环境的交互建模为马尔可夫决策过程:在执行动作后,智能体根据环境反馈获得奖励,并转入下一状态。而在本文提出的事件驱动的 MDP 中,状态转移是由入侵检测的输出驱动的。在此基础上,DQN 通过 Bellman 方程迭代逼近最优 Q 值函数[19]:

$$Q(\boldsymbol{s}_t,A_t)=R_t+\gamma\max_{A'}Q(\boldsymbol{s}_{t+1},A')。\tag{7}$$

式中:γ 为折扣因子。为稳定并高效地学习 Q 函数,本文采用目标网络加经验回放的标准 DQN 训练流程。首先,将环境交互过程中产生的四元组$(\boldsymbol{s}_t,A_t,R_t,\boldsymbol{s}_{t+1})$依次存入经验池 H。随后,从经验池中随机采样一个大小为 W 的小批量样本,并记第 i 个样本为$(\boldsymbol{s}_i,A_i,R_i,\boldsymbol{s}_i')$,其中 $\boldsymbol{s}_i$ 与 $\boldsymbol{s}_i'$ 分别由网络状态 S_t 和 S_{t+1} 按式(3)编码得到。基于该小批量样本,可按式(8)构造单步目标值

$$y_i=R_i+\gamma\max_{A'}Q_{\theta^-}(\boldsymbol{s}_i',A')。\tag{8}$$

式中:θ^-为目标网络参数。基于该小批量样本,采用均方误差损失函数度量在线 Q 网络 Q_θ 输出与目标值之间的偏差,即

$$L(\theta)=\frac{1}{W}\sum_{i=1}^{W}(y_i-Q_\theta(\boldsymbol{s}_i,A_i))^2。\tag{9}$$

式中:W 为批量大小;i 为样本编号。执行阶段采用 ε-greedy 策略进行探索-利用权衡。算法流程如算法 2 所示,在该方法中,DQN 的作用并非直接改变具体防御机制,而是依据网络状态在候选 MTD 动作之间进行动态选择,从而提高策略部署的针对性。由于状态输入同时反映了安全事件类别、当前防御配置以及网络服务时延影响,DQN 能够更好地区分不同状态下各策略的适用性。因此该方法能够在防御收益与网络服务时延影响之间取得更合理的平衡。

算法 2 基于 DQN 的 MTD 策略自适应选择算法。

输入:初始状态 S_0、在线 Q 网络 Q_θ、目标 Q 网络 Q_{θ^-}、经验池 H、训练轮数 M、每轮最大步数 T、小批量大小 W、折扣因子 γ、探索率 ε、目标网络更新周期 P;

输出:训练后的在线 Q 网络 Q_θ。

① 初始化 Q_θ、Q_{θ^-}、H,并令 $p=0$;

② for *episode*=1 to M do

③ 由状态判定模块获得当前回合起始网络状态 S_0,并按式(3)编码为 DQN 输入向量 $\boldsymbol{s}_0$;

④ $\boldsymbol{s}\leftarrow\boldsymbol{s}_0$;

⑤ for step=0 to $T-1$ do

⑥ 根据 ε-greedy 策略选取 A_t;

⑦　执行动作 A_t,获得奖励 R_t 及下一网络状态 S_{t+1};

⑧　按式(3)将 S_{t+1} 编码为下一状态输入 $\boldsymbol{s}'$;

⑨　将($\boldsymbol{s}$,A_t,R_t,$\boldsymbol{s}'$) 存入经验池 H;

⑩　if $|H| \geq W$ then

⑪　随机采样一个小批量;

⑫　根据式(8)计算目标值 y_i;

⑬　根据式(9)计算误差函数 $L(\theta)$,更新 θ;

⑭　若 $p>0$ 且 $\mathrm{mod}(p, P)=0$,则令 $\theta^- \leftarrow \theta$;

⑮　end if

⑯　$\boldsymbol{s} \leftarrow \boldsymbol{s}'$;

⑰　$p \leftarrow p+1$

⑱　end for

⑲ end for

5　实验结果与分析

5.1　实验环境与参数设置

为验证 MSD-MTD 的有效性,本文从检测、防御与网络服务时延影响 3 个方面开展仿真评估。实验环境基于 Mininet-WiFi 2.4.3 构建,采用 Ryu 4.34 作为 SDN 控制器,负责流表下发与控制消息处理。网络拓扑基于 NetworkX 3.4.2 中的 Waxman 模型生成,其中 $\alpha_{\mathrm{Waxman}}=0.5$,$\beta_{\mathrm{Waxman}}=0.4$。所构建网络包含 1 个 SDN 控制器、4 台 OpenFlow 交换机和 9 台孪生应用节点,用于模拟基于 SDN 的数字孪生网络场景[20]。该拓扑在保证网络规模可控的同时,能够较好地体现节点间连接的随机性与层次性,为后续策略评估提供统一的实验基础。为保证实验过程的可复现性,本文在 Mininet-WiFi 中采用动态链路控制方式实现欺骗节点的启用与停用。初始化阶段,除真实服务节点外,系统预先创建一组不承载真实业务的欺骗主机;运行过程中,通过控制欺骗节点与交换机之间链路的启停,模拟欺骗节点数量的动态增减。

入侵检测模块采用 ModernTCN 实现,其结构参数采用分类任务的默认设置,不再单独展开。策略学习基于 DQN 实现,优化器采用 Adam。DQN 训练与实验关键参数设置见表 1。DQN 的关键参数参考文献[21]设置,例如学习率为 0.001,批大小为 64,目标网络更新周期为 320。

5.2　网络状态感知能力

基于 3 个现实数据集评估 MSD-MTD 的网络状态感知能力:CICIDS-2017[22]、CSE-CIC-IDS2018[23](简称 CICIDS-2018)以及 UNSW-NB15[24]。本文重点关注 Benign、DoS/DDoS 与 Reconnaissance 这 3 类安全事件。其中,CICIDS-2017 与 CICIDS-2018 仅包含 Benign 和 DoS/DDoS 两类样本,UNSW-NB15 包含 Benign、DoS 和 Reconnaissance 这 3 类样本,并按 7:3的比例进行分层划分,构建训练集与测试集。为更贴近防御决策场景,实验剔除了与目标攻击类别无关的时段,仅保留攻击阶段内的正常流量与攻击流量片段。

表 1　DQN 训练与实验关键参数设置

Table 1　Key parameters for DQN training and experiments

参数	设置
训练轮数 M	5 000
每轮最大步数 T	50
小批量大小 W	64
学习率	0.001
折扣因子 γ	0.9
目标网络更新周期 P	320
隐藏层维度	[128,128,128,128]
随机种子	2 021, 2 022, 2 023, 2 024, 2 025

考虑到本文关注的是状态判定对后续防御策略的支撑作用,故主要采用整体准确率与各类识别率评价检测结果。评价指标以准确率为主。整体准确率定义为

$$Acc=\frac{TP+TN}{TP+TN+FP+FN}。\tag{10}$$

式中:TP 为将某类恶意事件正确判为该类的样本数;TN 为将非该类样本正确判为其他类的样本数;FP 与 FN 分别为误报与漏报的样本数。为更直观反映各事件类型的识别能力,本文给出各类事件的识别率,即该类真实样本中被正确识别的比例:

$$Acc_c=\frac{TP_c}{TP_c+FN_c}。\tag{11}$$

式中:c 为具体事件类型。

表 2 为各数据集安全事件识别结果。

表 2　各数据集安全事件识别结果

Table 2　Identification results of safety events in each dataset

数据集	事件类型	识别率/%
CICIDS-2017	Benign	95.63
	DoS/DDoS	96.73
CICIDS-2018	Benign	97.14
	DoS/DDoS	94.07
UNSW-NB15	Benign	97.04
	Reconnaissance	85.20
	DoS	95.13

从表 2 结果可知,在单类指标中,DoS/DDoS 流量识别效果通常优于 Reconnaissance 流量。这主要是因为 DoS/DDoS 流量在持续时间与流量形态上更易形成稳定模式,而 Reconnaissance 流量分布更为分散,且部分低频 Benign 流量可能对其造成干扰。总体来看,MSD-MTD 的检测模块能够较稳定地识别主要安全事件,为后续状态判定与策略选择提供输入。与此同时,不同数据集上的结果也表明,检测性能不仅与攻击类型有关,还会受到数据集流量构成与样本分布差异的影响,但整体结论保持一致,即所构建方法能够为后续防御决策提供较为可靠的状态感知基础。

5.3 消融实验与收敛性能

5.3.1 实验配置与评价指标

为量化 MSD-MTD 各组成模块的贡献,并区分协同欺骗与 DQN 决策带来的提升,本文设置 4 种对比配置:M(仅采用突变)、S(仅采用服务欺骗)、M+S(采用突变与服务欺骗协同)以及 M+S+DQN(本文方法)。其中,M 仅采用 IP 地址突变和服务端口突变,不启用服务欺骗节点;S 仅启用服务欺骗节点并关闭地址突变,欺骗节点数量固定;M+S 同时启用地址突变与服务欺骗,但不使用 DQN,欺骗节点数量按预设规则调整;M+S+DQN 在 M+S 基础上引入 DQN,实现欺骗节点数量与策略的自适应调节。

为保证对比公平,4 种配置采用相同的网络拓扑、流量数据集、时间窗划分、入侵检测模型与参数设置。对于学习型配置 M+S+DQN,训练阶段共进行 5 000 轮,每轮包含 50 步;训练完成后冻结策略参数,并在 5 个不同随机种子下分别评估。对于非学习型配置 M、S 和 M+S,无须训练,直接在相同实验设置下重复运行 5 次,结果取均值。

采用防御成功率(DSR)作为核心评价指标,定义为

$$DSR = \left(1 - \frac{\sum_{k=1}^{n} h_k}{\sum_{k=1}^{n} H_k}\right) \times 100\%。 \tag{12}$$

式中:n 为节点数量;h_k 为到达目标节点的恶意流量;H_k 为节点 k 的恶意流量总量。计算单次评估回合的 DSR,结果取每 100 个评估回合的平均值。

5.3.2 收敛性能

图 6 给出了 3 个数据集上的收敛曲线,为 5 次独立运行结果的均值。CICIDS-2017 与 CICIDS-2018 上的防御成功率在训练早期提升较快,随后逐步趋于稳定;UNSW-NB15 上的提升过程相对较缓,但最终也达到较稳定水平。结果表明,MSD-MTD 在不同数据集上均具有较好的训练稳定性与收敛性。

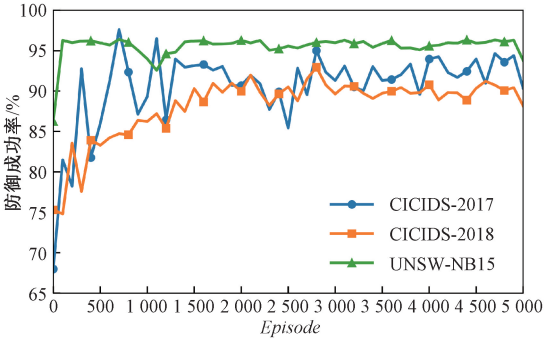


图 6 3 个数据集下 MSD-MTD 的训练收敛曲线

Figure 6 Training convergence curves of MSD-MTD on three datasets

5.3.3 消融结果

在相同网络拓扑、攻击流量与参数设置下,本文对 4 种配置进行消融评估。表 3 中的 DSR 取 5 次独立运行的均值。由表 3 可知,M+S 在各数据集上的 DSR 均优于仅采用突变或仅采用服务欺骗的配置,说明服务欺骗能够在突变触发间隙持续承接可疑流量,从而有效弥补纯突变策略在防御连续性上的不足。进一步引入 DQN 后,M+S+DQN 在 3 个数据集上的 DSR 与 M+S 总体相当,其中在 CICIDS-2017 上略有提升,在 CICIDS-2018 和 UNSW-NB15 上差异较小。这表明,在当前实验设置下,引入 DQN 后方法能够保持较稳定的防御效果,并在部分场景下表现出一定的性能提升。

表 3 不同消融配置下的 *DSR* 对比

Table 3 Comparison of *DSR* with different ablation configurations

数据集	*DSR*/%			
	M	S	M+S	M+S+DQN
CICIDS-2017	74.85	78.42	92.39	93.36
CICIDS-2018	82.72	84.15	88.30	88.20
UNSW-NB15	89.88	91.45	95.58	95.50

5.4 防御性能分析

将本文方法 MSD-MTD 与基线方法进行对比。所有方法均在相同网络拓扑与流量设置下评估,评估阶段运行固定轮次,DSR 取 5 次实验均值。基线方法介绍如下。

(1)IP Hopping。参考 IP hopping 类移动目标防御的基本思想[25],在 SDN 框架下对主机虚拟 IP 进行周期性突变,并在短暂过渡期保留旧映射以保障会话连续。为保证对比公平并隔离突变动作本身的贡献,该基线不引入额外的在线检测分类器与自适应频率切换机制,其启用的防御动作集合与消融

实验中的 M 配置一致,用于代表“仅依赖地址突变”的典型防护方式。

(2)ID-HAM[14]。在 SDN 中将主机地址突变表示为 MDP,采用优势 Actor-Critic 学习扫描行为,借助可满足性模理论(SMT)约束生成可行的地址块分配,并以滑动窗口保障会话连续,实现自适应地址块重分配与周期性虚拟 IP 地址突变。

图 7 展示了在不同数据集上各方法的防御性能对比。总体来看,MSD-MTD 在 3 个数据集上都取得了最高的 *DSR*,且波动相对更小,说明其防御效果不仅更好,也更稳定。从 3 组结果看,IP Hopping 和 ID-HAM 虽然都能通过地址变化削弱攻击者对目标的持续定位,但在突变触发间隙,攻击者仍可能利用已有侦察结果继续推进攻击,因此 *DSR* 提升有限。相比之下,MSD-MTD 在突变策略基础上进一步引入服务欺骗,在间隔期内将可疑流量重定向至欺骗节点进行承接与隔离,使攻击者更难继续利用上一轮侦察信息,从而有效弥补防御空窗。

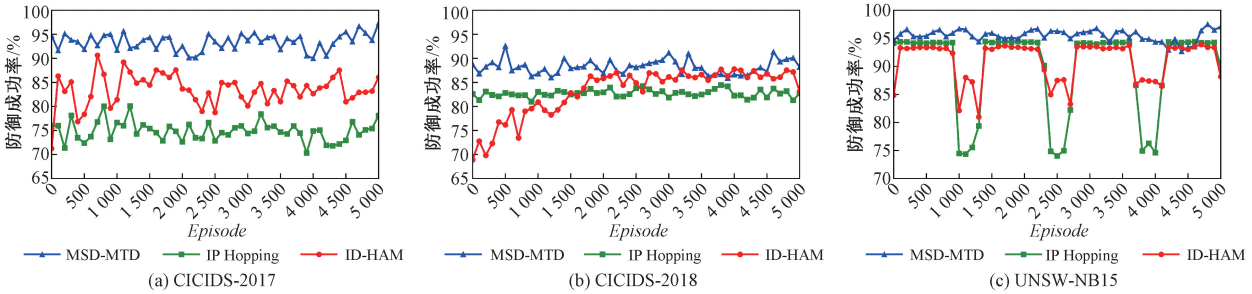


图 7　不同数据集下各方法的 *DSR* 对比

Figure 7　Comparison of *DSR* of different methods on three datasets

从平均 *DSR* 结果来看,MSD-MTD 在 CICIDS-2017、CICIDS-2018 和 UNSW-NB15 上的取值分别为 93.36%、88.20% 和 95.50%。相较于 IP Hopping,MSD-MTD 分别提升 18.51 百分点、5.48 百分点和 5.64 百分点;相较于 ID-HAM,分别提升 9.94 百分点、5.07 百分点和 4.34 百分点。虽然在不同数据集上的提升幅度有所不同,但 3 组实验结果均表明,MSD-MTD 能够取得更好的防御效果。

5.5　防御对网络服务影响分析

利用 *RTT* 评估不同防御方案对网络服务的影响。*RTT* 代表数据包完成一次端到端传输及确认返回的耗时,可综合反映路径切换、连接重建、流表更新与控制器下发等操作带来的时延波动。本文对前 300 s 内各方案的 *RTT* 进行对比,结果如图 8 所示。

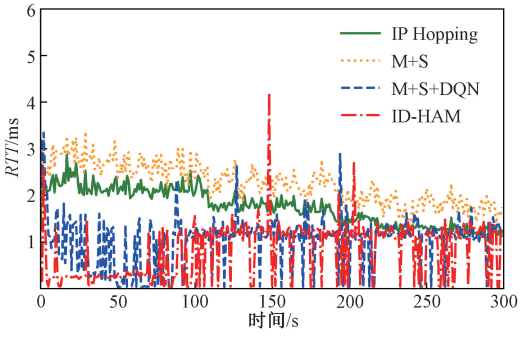


图 8　不同防御配置下的 *RTT* 对比

Figure 8　*RTT* comparison with different defense configurations

总体来看,M+S+DQN 的 *RTT* 主要分布在 0~2 ms 之间,与 ID-HAM 基本接近;M+S 的 *RTT* 主要分布在 2~4 ms 之间;IP Hopping 平均为 2 ms。其原因在于 M+S+DQN 与 ID-HAM 可按需触发配置调整,避免频繁、无差别的周期性重配置开销;而 IP Hopping 和 M+S 在固定周期突变下更容易引入额外时延。综上,本文所提方法在保持较低 *RTT* 的同时获得更优防御效果,对网络服务的影响更有限。

6　结论

本文面向数字孪生网络提出了 MSD-MTD 方法,通过协同地址突变、服务端口突变与服务欺骗缓解突变策略触发间隙的防御空窗,并结合跨节点流量对齐、特征选择、ModernTCN 状态识别与 DQN 策略选择,实现网络状态感知和 MTD 策略自适应部署。仿真结果表明,在本文设定的实验条件下,MSD-MTD 在 3 个数据集上均取得了更优的平均防御成功率,且 *RTT* 主要分布在 0~2 ms 之间,说明该方法在一定程度上兼顾了防御效果与网络服务时延影响。

尽管如此,当前研究仍主要基于受控仿真环境,对复杂强对抗场景和真实大规模数字孪生网络的验证仍显不足。后续工作将围绕复杂攻击场景下的持续防护、更多类型 MTD 策略的联合调度以及原型系统验证展开研究。

参考文献:

[1]　Lin Xingqin, Kundu L, Dick C, et al. 6G digital twin networks: from theory to practice[J]. IEEE Communica-

tions Magazine, 2023, 61(11): 72-78.
[2] Nguyen H X, Trestian R, To D, et al. Digital twin for 5G and beyond[J]. IEEE Communications Magazine, 2021, 59(2): 10-15.
[3] Alcaraz C, Lopez J. Digital twin: a comprehensive survey of security threats[J]. IEEE Communications Surveys & Tutorials, 2022, 24(3): 1475-1503.
[4] Wang Weizheng, Yang Yaoqi, Khan L U, et al. Digital twin for wireless networks: security attacks and solutions [J]. IEEE Wireless Communications, 2024, 31(3): 278-285.
[5] He Ke, Kim D D, Asghar M R. Adversarial machine learning for network intrusion detection systems: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 538-566.
[6] Lei Cheng, Zhang Hongqi, Tan Jinglei, et al. Moving target defense techniques: a survey[J]. Security and Communication Networks, 2018, 2018: 3759626.
[7] Cho J H, Sharma D P, Alavizadeh H, et al. Toward proactive, adaptive defense: a survey on moving target defense[J]. IEEE Communications Surveys & Tutorials, 2020, 22(1): 709-745.
[8] Zhang Tao, Xu Changqiao, Lian Yibo, et al. When moving target defense meets attack prediction in digital twins: a convolutional and hierarchical reinforcement learning approach[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(10): 3293-3305.
[9] Rehman Z, Gondal I, Ge Mengmeng, et al. Proactive defense mechanism: enhancing IoT security through diversity-based moving target defense and cyber deception [J]. Computers & Security, 2024, 139: 103685.
[10] Masud M T, Keshk M, Moustafa N, et al. Vulnerability defence using hybrid moving target defence in Internet of Things systems [J]. Computers & Security, 2025, 153: 104380.
[11] Zhou Yuyang, Cheng Guang, Ouyang Zhi, et al. Resource-efficient low-rate DDoS mitigation with moving target defense in edge clouds[J]. IEEE Transactions on Network and Service Management, 2025, 22(1): 168-186.
[12] Hu Hongchao, Zhang Shuaipu, Cheng Guozhen, et al. ReDoS defense method based on moving target defense in cloud-native environment[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(2): 72-79. [扈红超, 张帅普, 程国振, 等. 云原生环境下基于移动目标防御的 ReDoS 防御方法[J]. 郑州大学学报(工学版), 2024, 45(2): 72-79.]
[13] Tan Jinglei, Jin Hui, Zhang Hongqi, et al. A survey: when moving target defense meets game theory[J]. Computer Science Review, 2023, 48: 100544.
[14] Zhang Tao, Xu Changqiao, Shen Jiahao, et al. How to disturb network reconnaissance: a moving target defense approach based on deep reinforcement learning[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 5735-5748.
[15] Beltrán-López P, Pérez M G, Nespoli P. Cyber deception: taxonomy, state of the art, frameworks, trends, and open challenges[J]. IEEE Communications Surveys & Tutorials, 2026, 28: 1520-1556.
[16] Pai V, Pai K, Manjunatha S, et al. Adaptive network anomaly detection using machine learning approaches [J]. EURASIP Journal on Information Security, 2025, 2025: 29.
[17] Luo Donghao, Wang Xue. ModernTCN: a modern pure convolution structure for general time series analysis[C]// 12th International Conference on Learning Representations. Appleton: ICLR, 2024: 1-43.
[18] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518: 529-533.
[19] Cui Mingxiu. DQN and dynamic feedback for multitask scheduling optimization in engineering management[J]. International Journal of Low-Carbon Technologies, 2024, 19: 2279-2286.
[20] Kumar P, Kumar R, Aljuhani A, et al. Digital twin-dri-ven SDN for smart grid: a deep learning integrated blockchain for cybersecurity[J]. Solar Energy, 2023, 263: 111921.
[21] Li Qiuxiang, Wu Jianping. Optimizing the effectiveness of moving target defense in a probabilistic attack graph: a deep reinforcement learning approach[J]. Electronics, 2024, 13(19): 3855.
[22] Sharafaldin I, Lashkari A H, Ghorbani A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]// 4th International Conference on Information Systems Security and Privacy. Cham: Springer, 2018: 108-116.
[23] Amazon Web Services. A realistic cyber defense dataset (CSE-CIC-IDS2018)[DS/OL]. [2026-03-19]. https://registry.opendata.aws/cse-cic-ids2018/.
[24] Moustafa N, Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS). Piscataway: IEEE, 2015: 1-6.
[25] Xu Xiaoyu, Hu Hao, Liu Yuling, et al. An adaptive IP hopping approach for moving target defense using a lightweight CNN detector[J]. Security and Communication Networks, 2021, 2021: 8848473.

(下转第133页)