基于深度确定性策略梯度的随机路由防御方法

徐潇雨[1,2]，胡浩[1,2]，张红旗[1,2]，刘玉岭[3]

（1. 信息工程大学密码工程学院，河南 郑州 450001；
2. 河南省信息安全重点实验室，河南 郑州 450001；3. 中国科学院信息工程研究所，北京 100190）

摘 要：针对现有随机路由防御方法对数据流拆分粒度过粗、对合法的服务质量（QoS）保障效果不佳、对抗窃听攻击的安全性有待提升等问题，提出一种基于深度确定性策略梯度（DDPG）的随机路由防御方法。通过带内网络遥测（INT）技术实时监测并获取网络状态；通过 DDPG 方法生成兼顾安全性和 QoS 需求的随机路由方案；通过 P4 框架下的可编程交换机执行随机路由方案，实现了数据包级粒度的随机路由防御。实验表明，与其他典型的随机路由方法相比，所提方法在对抗窃听攻击中的安全性和对网络整体 QoS 的保障效果均有提升。

关键词：随机路由；深度确定性策略梯度；窃听攻击；移动目标防御

中图分类号：TP939

文献标识码：A

DOI: 10.11959/j.issn.1000−436x.2021093

Random routing defense method based on deep deterministic policy gradient

XU Xiaoyu[1,2], HU Hao[1,2], ZHANG Hongqi[1,2], LIU Yuling[3]

1. Cryptography Engineering Institute, Information Engineering University, Zhengzhou 450001, China
2. Henan Key Laboratory of Information Security, Zhengzhou 450001, China
3. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100190, China

Abstract: To solve the problem of the existing routing shuffling defenses, such as too coarse data flow splitting granularity, poor protection effect on legitimate QoS and the security against eavesdropping attacks needed to be improved, a random routing defense method based on DDPG was proposed. INT was used to monitor and obtain the network state in real time, DDPG algorithm was used to generate random routing scheme considering both security and QoS requirements, random routing scheme was implemented with programmable switch under P4 framework to realize real-time routing shuffling with packet level granularity. Experiment results show that compared with other typical routing shuffling defense methods, the security and QoS protection effect of the proposed method are improved.

Keywords: random routing, deep deterministic policy gradient, eavesdropping attack, moving target defense

1 引言

随着信息技术的发展，网络在惠及诸多领域的同时，其安全性也面临严峻挑战。赛门铁克公司 2019 年发布的互联网安全研究报告[1]指出，目标性攻击已成为当前网络犯罪的主要方式，最活跃团伙在过去三年中攻击的企业数平均达到了 55 家。

常规网络的静态属性使攻击者容易捕捉攻击目标并发动长期有效的攻击。移动目标防御[2]是美国研究人员针对当前网络安全博弈中防御方所处的弱势地位提出的革命性技术，其旨在“改变游戏规则”，打破攻守不平衡的现状。移动目标防御通过不断改变网络的关键属性，使攻击方可能利用的攻击面不断发生变化，从而达到迷惑攻击方、增加

收稿日期：2021−01−13；修回日期：2021−03−31
通信作者：胡浩，wjjhh_908@163.com
基金项目：国家自然科学基金资助项目（No.61902427, No.61802404）
Foundation Item: The National Natural Science Foundation of China (No.61902427, No.61802404)

攻击难度和成本、降低攻击收益的目的。

路由路径是网络攻击面的重要组成部分。针对固定的转发路径，攻击者可发动窃听攻击[3]、黑洞攻击[4]、拒绝服务攻击[5]等，对网络正常业务构成严重威胁。随机路由防御技术通过动态地改变通信双方的路由路径，以规避恶意窃听等攻击行为。因此，随机路由防御技术已成为移动目标防御下的一项重要技术和热点研究方向。

本文提出一种基于深度确定性策略梯度（DDPG, deep deterministic policy gradient）[6]的随机路由防御方法。首先通过带内网络遥测（INT, in-band network telemetry）技术获取实时的网络状态；然后将获取的实时网络状态输入 DDPG 方法，依据实时网络状态和攻击者行为特性生成符合安全性和服务质量（QoS, quality of service）需求的随机路由方案；最后基于 P4[7]网络架构执行该随机路由方案，实现了数据包级粒度的随机路由。

2 背景知识与相关工作

2.1 移动目标防御基本原理

移动目标防御技术通过变换网络要素信息，例如路由路径等，切断攻击方的网络侦察、目标攻击。移动目标防御迫使攻击者不断追逐攻击目标，增大攻击方成本，消除攻击方的时间优势和信息不对称优势。

采用移动目标防御技术的网络系统依据安全需求和 QoS 需求生成新的跳变要素，并将新的跳变要素更新至当前系统，如图 1 所示。其中，跳变要素由攻击属性决定，例如，跳变要素为路由路径则可以有效减少窃听攻击；安全需求也由攻击属性决定，同样针对窃听攻击，安全需求是使数据包在传输中尽可能不被攻击方截获；QoS 需求由网络状态决定，例如，FTP 要求路由路径有充足的带宽保证，HTTP 对路由路径的时延有较高要求。

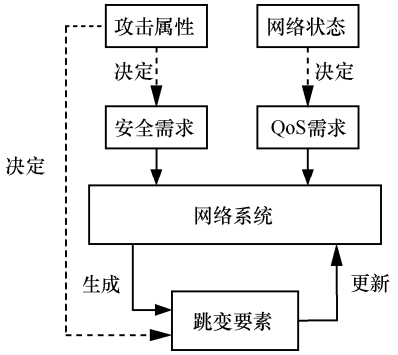


图 1 移动目标防御基本原理

2.2 随机路由防御相关工作

Duan 等[8]首先提出了一种高效的随机路由方法，被称为随机路由跳变（RRM, random route mutation），利用可满足性模理论（SMT, satisfiability modulo theory）求解满足要求的路由路径。Jafarian 等[9]考虑了链路安全和防御效益，使用博弈论来探讨路由方案生成。Zhao 等[10]提出了一种双跳变通信（DHC, double-hop communication）方法，同时对端点信息和路由路径实施无规则的跳变。Aseeri 等[11]指出，在某些情况下，现有随机路由方法仍然会导致 100%的数据暴露，在此基础上提出对网络中的确认数据包也实施随机路由，以避免严重的数据泄露。Zhou 等[12]提出了一种时空随机优化路由跳变方法（RRO-RM, spatio-temporal stochastic optimization route mutation），确保安全性从流（用户）和节点（基础设施）2 个角度得到保证。Duan 等[13]提出了一种主动的路由跳变方法，在量化网络中实体脆弱性的基础上实施周期性的路由跳变。Liu 等[14]利用网络异常检测触发随机路由跳变（AT-RRM, anomaly triggered random routing mutation），使用改进的蚁群算法求解可行的路由方案，进一步增加窃听攻击的难度。Lei 等[15]提出一种最优路径跳变方法，通过基于安全容量矩阵的最优路径跳变生成方法选取最优跳变路径和跳变周期组合，以实现防御收益的最大化。Zhang 等[16]提出了一种安全感知 Q-learning 算法，从路由跳变空间中迭代选择路径，并进行安全感知自适应调整学习速率，同时，从理论上证明了算法的最优收敛性。Zhang 等[17]进一步提出了一种适用于大规模状态−动作空间的基于深度强化学习的路由跳变方案，验证了该方法在防御性能和收敛速度上较已有方法有较大的提高。

当前，随机路由防御已成为移动目标防御下的研究热点。然而，分析现有方法发现，其仍存在以下问题。

1) 对数据流的随机拆分粒度过粗。已有方法通过将数据流拆分为多个“子流”，为不同“子流”分配不同路由路径。然而，“子流”中的数据包是连续的，攻击者一旦截获某个或多个子流，可能从连续数据中解析出有价值的信息。

2) 对网络合法 QoS 的保障效果不够理想。一方面，不同网络应用对网络资源的需求是不同的，已有方法对不同应用的不同需求未做区分，导致网络整体 QoS 保障质量不佳；另一方面，由于对网络

状态感知的实时性、准确性、全面性不够，生成路由方案时考虑的 QoS 因素不够准确，导致 QoS 保障质量不佳。

3) 对复杂随机路由问题的求解能力不足。既有的可满足性理论、蚁群算法等面对高维数据处理能力不足、搜索空间大等缺点和高复杂性路由路径生成问题时，表现出求解能力不足，进而造成对抗窃听攻击时防御效果有待提高。

3　威胁模型

本文以链路窃听攻击建立威胁模型。常见的窃听攻击会潜入某些节点（如交换机端口）以窃听特定的链路。然而，窃听攻击者的攻击行为不仅表现为窃听行为本身，还表现为通过蠕虫传播等方式有目的地在网络节点之间转移，以及获取网络拓扑、分析可能的路由路径等。随机路由防御方法的设计需要考虑到窃听攻击者的应变能力。

网络拓扑可被建模为有向图，E 表示图中的边集合，即网络中连接节点的链路。

3.1　攻击机理分析

链路窃听攻击的实施过程分为 2 个阶段：初始阶段，攻击方通过社会工程等方式将窃听恶意程序部署至网络中，以窃听一定数量的链路；后续阶段，主动或被动地将已有的窃听恶意程序转移到其他节点上，以窃听其他链路。在窃听点转移过程中，假设攻击方具有以下能力。

3.1.1　路由路径覆盖

由于攻击者熟知当前网络的拓扑结构，当已潜伏的窃听恶意软件截获到某数据包时，可从中读取源和目的地址，并依据网络拓扑计算出其他同流数据包的可能路由路径。攻击者采用简单搜索方法（SSM, simple search method）[8]计算可能的路由路径，形式化表示为

$$R = \mathrm{SSM}(\mathrm{Pkt}) \tag{1}$$

其中，Pkt 表示截获的数据包，R 表示可能的路由路径集合。设当前正在窃听的路由路径集合为 C，窃听点转移将侧重选择集合 $E - R \wedge C$ 中的链路，即当前未覆盖的路由路径所包含的链路。

3.1.2　重点链路转移

攻击者在转移窃听点时，更倾向于能截获更多数据包的链路。当转移发生时，攻击者会记录在原窃听链路上截获的数据包数量，构成二元组 $\{l_k, n_k\}$。其中，l_k 表示一条链路，n_k 表示截获数据包数量。窃听点转移将侧重选择当前 n_k 值更大的链路。

3.1.3　检测规避

对攻击者而言，网络中的恶意软件检测是客观存在的。攻击者在转移窃听点时，更倾向于检测能力弱的链路，以降低被检测移除的可能性。当某窃听点发生被动转移（即被检测程序移除）时，攻击者会记录恶意程序在原窃听链路上的生存时间，构成二元组 $\{l_k, t_k\}$。其中，t_k 表示最近一次记录在链路 l_k 上的生存时间。窃听点转移将侧重选择当前 t_k 值更大的链路，该项能力的优先级次于重点链路转移。

3.2　攻击行为约束

攻击方实施窃听攻击的行为符合两项约束，即时间约束和空间约束。时间约束限制攻击者在一段链路上窃听的时间；空间约束限制窃听的链路数量占整个网络链路数量的比例。

3.2.1　时间约束

网络中的入侵检测系统有能力发现窃听攻击的存在并清除寄生的恶意软件。假设某条链路的恶意软件被移除的概率随着寄生时间的增长而增大。由于入侵检测系统非本文研究内容，以式(2)模拟入侵检测系统清除恶意软件。

$$P_C = \frac{e^{t_C}}{k_C} \tag{2}$$

其中，P_C 表示恶意软件被清除的概率，t_C 表示恶意软件寄生的时长，k_C 表示既定常数。

3.2.2　空间约束

攻击者为了尽可能隐藏窃听行为，必须控制任意时刻窃听的链路数量。约定攻击在任意时刻窃听的链路数量占整个网络链路数量的比例在一个既定的阈值以下，如式(3)所示。

$$\frac{\mathrm{card}(C)}{\mathrm{card}(E)} \leqslant T_C \tag{3}$$

其中，C 表示当前攻击者窃听的链路集合，T_C 表示既定阈值，函数 $\mathrm{card}(\cdot)$ 表示求集合元素的数量。

3.3　攻击目标

攻击者意图窃听尽可能多的全网通信数据包，以指标 PPE 衡量攻击者的窃听收益，如式(4)所示。

$$\mathrm{PPE} = \frac{N^E}{N^P} \tag{4}$$

其中，N^E 表示截获的数据包数量，N^P 表示全网传输的数据包总量。攻击方目标是使 PPE 值尽可能提高。

4　基于 DDPG 的随机路由方案

本文在 SDN（software defined network）下生成和实施随机路由方案，架构如图 2 所示。利用 SDN 数控分离的特点，使用 INT 技术实时监测网络状态。位于控制平面的 DDPG 方法依据实时网络状态，生成随机路由方案。生成的随机路由方案被动态下载至数据平面而被执行。主要系统参数与含义如表 1 所示。

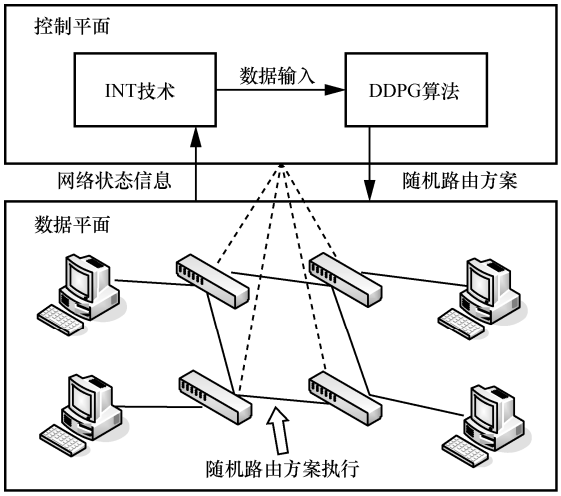


图 2　SDN 随机路由架构

表 1　　系统参数与含义

参数	含义
F	网络中的一条数据流
RRS	数据流 F 的随机路由方案
Vol^t	第 t 防御周期中传输的流量大小
N^t_{dsp}	第 t 防御周期中传输的对时延敏感的数据包总量
$\mathrm{Td}^t_{\mathrm{dsp}}$	第 t 防御周期中传输数据包的总时延
S_t	第 t 防御周期获取的状态（Sate）
A_t	第 t 防御周期执行的动作（Action）
R_t	第 t 防御周期获得的奖励（Reward）
Th	差异性约束单个门限值
th	差异性约束总体门限值

4.1　INT 网络状态监测

本文提出的 DDPG 方法以实时的网络为学习环境。DDPG 方法需要依据当前的网络状态信息，生成新的随机路由方案，采用 INT 技术[18]获取实时的网络状态，为 DDPG 方法做数据准备。

应用 INT 技术，网络中的交换机可对其转发的数据包插入 INT 头部和 INT 元数据。INT 头部声明了 INT 元数据的内容和格式，INT 元数据则包含交换机的内部信息，即当前的网络状态。

INT 收集的交换机内部信息包括交换机 ID、当前数据包队列长度、当前可用带宽和当前排队等待时间（即数据包到达交换机至被转发地时间长度）。交换机 ID 用于统计不同路由路径的流量。当前数据包队列长度、当前可用带宽和当前排队等待时间作为 QoS 因素，输入 DDPG 方法使生成 RRS 时，可在确保安全性的前提下兼顾网络 QoS。受保护网络中的交换机均可执行该 INT 处理，在面临首次生成 RRS 和更新 RRS 时对数据包插入 INT 数据。

4.2　面向随机路由的 DDPG 方法

以 INT 技术输出的实时网络状态为数据基础，使用 DDPG 方法生成随机路由方案。DDPG 是一种典型的深度强化学习算法，具有可处理高维数据、学习效率高、模型容量大等优势，能够应用于复杂随机路由问题的求解。

4.2.1　智能体结构

DDPG 是一种基于“演员–评论家”模式的强化学习算法。“演员”和“评论家”组成智能体（Agent）。“演员”用于拟合策略函数，其输入为环境的当前状态（State，即 INT 采集的实时网络数据），输出为行为（Action，即随机路由方案）；“评论家”用于拟合价值函数，其输入为“演员”的行为、环境的变化状态和环境给予该行为的奖励（Reward），输出为对参与者行为的评价。“演员”将“评论家”的评价作为更新参数的梯度（Gradient），通过反向传播（BP, back propagation）算法更新自身参数。“演员”与环境的每一次互动都将被记录在样本库中。智能体每次学习时仅从样本库中抽取少量样本进行拟合。“演员”和“评论家”均采用 off-policy 方式训练，即“演员”和“评论家”均由一个在线网络和一个目标网络组成，在线网络学习数论后可用于更新目标网络。

在基于 DDPG 的随机路由方案生成中，“演员”和“评论家”分别由 2 个结构相同但参数不同的深层神经网络构成。“演员”由目标策略网络和在线策略网络构成，“评论家”由在线 Q 网络和目标 Q 网络构成。DDPG 方法将策略网络和价值网络（Q 网络）分开，以实现 off-policy 学习，因此“演员”和“评论家”均由 2 个网络构成。“演员”和“评论家”各自的在线网络在训练中即时地更迭参数，经过特定步数的学习后再将自身参数更新至目标网络。

“演员”和“评论家”均使用卷积神经网络。“演员”网络由 3 个卷积层构成，每个卷积层均采用池化（Pooling）和 ReLU 激活；“评论家”网络由 2 个卷积层和一个全连接层构成，卷积层同样采用池化（Pooling）和 ReLU 激活。各层超参数（如卷积核尺寸）依据输入和输出数据尺寸决定，即依据底层网络规模与可能的数据流数量决定。

4.2.2 算法流程

“演员”与环境互动一次称为一步（Step），“演员”和“评论家”各自的在线网络在每一步之后都会进行一次参数更新，而其各自的目标网络则会在固定步数之后将在线网络的参数复制到自身，以完成更新。一步即对应随机路由防御的一个防御周期。

与“演员”互动的环境即 SDN 的数据平面，其通过 INT 技术将实时的网络状态交付给“演员”。DDPG 方法则部署在 SDN 的控制平面。经过足够多步的互动后，经训练的 DDPG 智能体将可以产生满足安全性和 QoS 需求的 RRS。

DDPG 方法的流程如算法 1 所示，其中，S 表示环境输出的状态，A 表示“演员”做出的动作（即生成的 RRS），R 表示在 S 下做出 A 时环境给予的奖励，S' 表示在 S 下做出 A 时环境输出的下一个状态。$\pi_\theta(\cdot)$ 和 $\pi_{\theta'}(\cdot)$ 分别表示目标策略网络和在线策略网络的状态到动作的映射。在训练中，m 表示单批样本量，$j\in[1,m]$；(S_j,A_j,S_j',R_j) 表示 m 个样本中的第 j 个；y_j 表示第 j 个样本经目标 Q 网络输出的目标 Q 值。基于 DDPG 的随机路由方案生成如算法 1 所示。智能体由环境获取当前状态 S 后，由在线策略网络计算得到对应的动作 A，而后对环境实施动作 A 并得到新的状态 S' 和奖励 R，并将上述要素组成的四元组 $\{S,A,S',R\}$ 存至样本库 D。每次学习从样本库 D 中抽取 m 个样本，依据算法 1 中所述的损失函数更新在线策略网络和在线 Q 网络。每经过 f_C 次学习后更新目标策略网络和目标 Q 网络。

算法 1　基于 DDPG 的随机路由方案生成

输入　在线策略网络 $\pi_\theta(\cdot)$；目标策略网络 $\pi_{\theta'}(\cdot)$；在线 Q 网络 Q 及其参数 ω，目标 Q 网络 Q' 及其参数 ω'；衰减因子 η；更新系数 τ；训练单批次样本量 m；目标网络更新频率 f_c

输出　最优在线策略网络 $\pi_\theta(\cdot)$

随机初始化 θ，ω；$\theta'=\theta$；$\omega'=\omega$；

由环境获取状态 S；

for t from 1 to T

　　计算动作 $A=\pi_\theta(S)$；

　　执行动作 A 并获取新状态 S' 和奖励 R；

　　存储四元组 $\{S,A,S',R\}$ 至样本库 D；

　　更新 $S=S'$；

　　由样本库中随机选取 m 个样本 $\{S_j,A_j,S_{j+1},R_t\}$，$j=1,2,\cdots,m$；

　　计算目标 Q 值：$y_j=R_j+\eta Q'(S_j',\pi_{\theta'}(S_j'),\omega')$；

　　更新损失函数 ω：$\frac{1}{m}\sum_{j=1}^{m}(y_j-Q(S_j,A_j,\omega))^2$；

　　更新损失函数 θ：$-\frac{1}{m}\sum_{j=1}^{m}Q(S_j,A_j,\omega)$；

　　if $i\%f_c==1$，更新 θ' 和 ω'：

　　　　$\omega'=\tau\omega+(1-\tau)\omega'$；

　　　　$\theta'=\tau\theta+(1-\tau)\theta'$；

　　end if

end for

4.2.3 关键变量设计

1) 状态

状态是当前环境情况的形式化表达，是生成 RRS、训练“演员”和“评论家”的输入数据。在 INT 技术的辅助下，状态被表示为一个二维矩阵，其表达的信息实时、准确地反映了当前 SDN 数据平面的网络状态。

如图 3 所示，二维矩阵的第一维度（横向）表示网络中交换机的序列，第二维度（纵向）表示对应交换机的 INT 信息，即该矩阵中的每一列均表示一个交换机的当前 INT 信息。由于 INT 采集了 4 类信息包括：交换机 ID、当前数据包队列长度、当前可用带宽和当前排队等待时间，因此状态矩阵的尺寸为 $N\times4$，其中，N 为网络中的交换机数量。

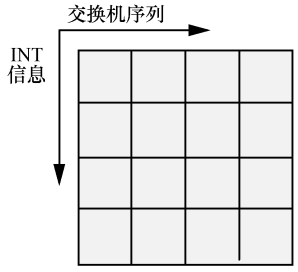


图 3　状态

2) 动作

动作即“演员”输出的一个全网 RRS。防御方为每一条可能的数据流生成单独的 RRS，即对每一条数据流指定经由它的每一条可行的路由路径转发的数据包数量比例。因此，“演员”输出的动作需要给出每一个可能的数据流指定 RRS。

所设计的动作变量被表示为一个三维矩阵，如图 4 所示。该矩阵的第一维度和第二维度共同指定数据流，第一维度指定一对主机排列（即指定源主机和目的主机），第二维度指定应用层协议。该矩阵的第三维度用于指定路由路径。

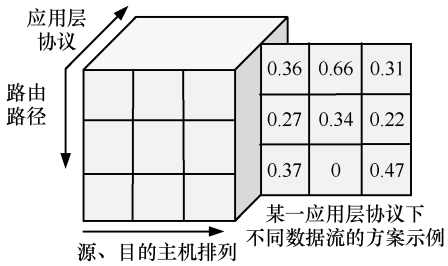


图 4 动作

图 4 中右侧的二维平面示例了某一应用层协议下不同数据流的随机路由方案。该二维平面的每一列均为一条数据流的 RRS，其指定了经由各个可行的路由路径转发的数据包数量比例，故每一列数值之和为 1（由“演员”网络的最后一个卷积层最后执行局部归一化实现）。由于不同数据流的可行路由路径数量不同，故该矩阵空缺部分使用“零填充”。

3) 奖励

奖励是对环境实施动作后得到的回馈，奖励值越大表明动作质量越高，反之表明动作质量越低。动作即一个时间周期内的 RRS，因此奖励应鼓励动作追求 2 个目标：安全性目标和 QoS 目标。第 t 步（防御周期）动作获得的奖励值的计算如式(5)所示。

$$R_t=\begin{cases}\mu\dfrac{\mathrm{Vol}^t}{T_{\mathrm{slot}}}-\gamma\dfrac{\mathrm{Td}^t_{\mathrm{dsp}}}{N^t_{\mathrm{dsp}}}, D_t>\mathrm{Th}, \forall d^t_F>\mathrm{th}\\ R_{\mathrm{BAD}}, \text{其他}\end{cases} \tag{5}$$

为鼓励追求安全性目标，应鼓励连续 2 个防御周期的 RRS 之间具有足够的差异性。式(5)中，d^t_F 表示数据流 F 的第 t 个 RRS 与第 $t-1$ 个 RRS 的欧氏距离，D_t 表示全网数据流的差异性平均值，Th 和 th 为固定阈值。令 $R_{\mathrm{BAD}}<0$，使生成的连续 2 个 RRS 差异性不足时，受到“惩罚”。

为鼓励追求 QoS 目标，需要兼顾网络的时延和带宽表现。以 Vol^t 表示在 t 防御周期中网络传输的流量大小，N^t_{dsp} 表示在 t 防御周期中网络中传输的对时延敏感的数据包数量（例如 HTTP 等应用的数据包），$\mathrm{Td}^t_{\mathrm{dsp}}$ 表示传输这些数据包的总时延，μ 和 γ 为人为设定的常数。平均时延越小、单位时间吞吐量越大，奖励值越高。

5 基于 P4 的细粒度随机路由实现

本文所提方法使用的 SDN 基于 P4 架构实现。使用 P4 架构的原因为，可利用 P4 架构对数据平面可编程、可定义数据平面与控制平面交互内容的特点，实现以下功能。

1) 支持 INT 获取实时网络状态。利用可编程交换机，人为定义交换机对数据包的处理过程。使在处理中将 INT 信息嵌入数据包，从而获取实时网络状态，并上传至控制平面。

2) 实现数据包级细粒度的随机路由。使数据流中的每个数据包的转发路径相互独立，不同数据包的转发路径之间无明显规律性。

5.1 工作原理

已有的基于 OpenFlow 协议的随机路由方案执行原理为：控制平面为数据流指定唯一的路由路径，并以流表的形式下发至数据平面的交换机，交换机按“流表”指定的唯一端口转发数据包。而后定期更新流表以实现路由路径的随机变化。该种方法是将数据流按序拆分为多个“子流”，每个“子流”由多个顺序连续的数据包组成。控制平面在“子流”的间隙下发新的流表，为后续“子流”切换新的路由路径。该种方法具有以下弊端。

1) 同一“子流”中的所有数据包使用相同的路由路径。部署在网络中特定位置的窃听程序可能截获一个或多个完整的“子流”。由于“子流”中的数据包的顺序是连续的，相比于顺序离散的数据包，顺序连续的数据包更容易被解析出有效信息。

2) 定期更新流表以实现路由路径的随机变化，需要控制平面和数据平面间频繁交互，产生了一定的通信数据量。

基于 P4 架构的数据包级粒度随机转发，可有

效缓解上述 2 个问题，其原理为：控制平面为数据流指定其所有可行的路由路径，即为数据平面的每个交换机同时下发多个可能被执行的流表。当数据包到达交换机时，交换机按一定概率为其随机选择一个可行的转发端口。在该方案中，控制平面仅需在最初下发流表时与数据平面交互一次即可。此外，由于转发端口的随机选择是以数据包级粒度执行的，可确保同一路径转发的数据包在顺序上是离散的，使窃听攻击程序难以截获顺序连续的数据包。数据包的路由路径仅在转发前即时决定，从而增加了防御行为的随机性和不可预测性。

“子流”级粒度与数据包级粒度随机转发对比如图 5 所示。在“子流”级粒度的随机转发中，若窃听攻击程序对交换机的某一端口实施窃听（如图 5 中端口 2），则可以截获完整的“子流”。而在数据包级粒度的随机转发中，其截获的数据包来自不同子流，被截获的数据包之间无连续性。

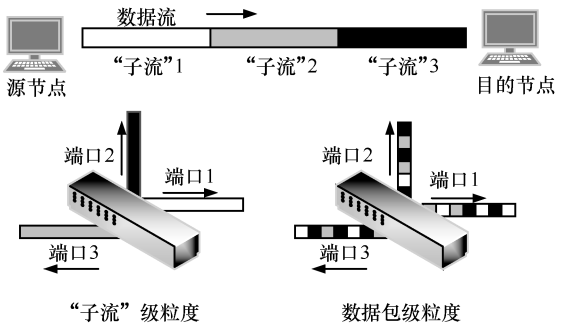


图 5　“子流”级粒度与数据包级粒度随机转发对比

针对“子流”级粒度随即转发可能造成的数据包失序问题，对于 TCP 及对数据包到达顺序敏感的 UDP 应用，例如 VoIP 等，采用改进的 FLARE 方法[19]缓解细粒度数据流拆分与数据包失序问题之间的矛盾，关键步骤如下。

1) 对于给定的交换机，当属于同一个数据流的任意 2 个数据包先后连续达到该交换机时，计算该 2 个数据包到达该交换机的时间差 $(t^{k+1}-t^{k})$。

2) 在可行路由路径集合中搜索，寻找满足条件的最大子集 R_{sub}，该子集中元素需满足条件：任意 2 个路由路径的当前时延差小于 1)中所述的时间差，如式(6)所示。

$$X=(t^{k+1}-t^{k})>\max(\text{Del}_{r_i}-\text{Del}_{r_j}),\ r_i,r_j\in R_{\text{sub}} \tag{6}$$

其中，Del_{r_i} 表示路由路径 r_i 的当前时延，要求 $X=\text{true}$。

3) 对于 1)中所述的 2 个数据包中后到达的数据包，仅可在 2)中得到的最大子集 R_{sub} 中随机选择路由路径，随机决定转发端口。

对于对数据包到达顺序敏感的网络流量，上述改进的 FLARE 方法虽然在数据流拆分粒度上做出了一定的妥协，却将在极大程度上缓和了数据包失序问题。而对于对数据包到达顺序不敏感的网络流量，例如 WebRTC 应用等，仍以数据包级粒度随机转发。

5.2　工作流程

当 DDPG 方法在控制平面生成 RRS 后，需要将 RRS 下发至数据平面，再由数据平面的可编程 P4 交换机执行 RRS。

5.2.1　随机路由方案下发

已知对于数据流 F，RRS 为其指定了经由各个可行的路由路径转发的数据包比例。将 RRS 部署至数据平面执行，须将随机路由方案映射为数据平面中各个交换机的执行方案，如式(7)所示。

$$\{\text{ES}_1,\text{ES}_2,\cdots,\text{ES}_N\}=M(\text{RRS}) \tag{7}$$

其中，ES_k 表示第 k 个交换机的执行方案。

对数据流 F 而言，数据平面中某交换机的执行方案可描述成为每一个可行的转发端口分配一个概率值，对于任意属于数据流 F 的数据包，将按所分配的概率随机决定转发端口。某交换机的执行方案如表 2 所示。

表 2　对数据流 F，某交换机的执行方案示意

转发端口	概率值
port_1	P_1
port_2	P_1
⋮	⋮
port_n	P_n

5.2.2　随机路由方案执行

P4 程序内涵转发端口随机选择功能，该程序被编译后装载至 P4 交换机。对于每一个到达 P4 交换机的数据包，P4 交换机为其随机选择一个可行的转发端口，具体方法为：在 P4 标准处理流程的 int_ingress 方法中，加入随机数生成函数和端口选择函数。使用随机数生成函数生成一个随机数，而后使用端口选择函数将产生的随机数映射至一个唯一的可行转发端口。在端口选择函数中，某端口

被选择的概率取决于 5.2.1 节中产生的交换机执行方案。

6 实验与分析

为了验证本文提出的基于 DDPG 的随机路由方法的可行性和防御效果，本节首先介绍包括软、硬件和算法超参数在内的实验设置，然后介绍实验的执行过程，最后从安全性和 QoS 这 2 个角度评价所提方法的性能表现。

6.1 实验设置

实验在 P4 架构下实施，SDN 的控制平面采用 P4 runtime，数据平面由支持 P4 语言的可编程交换机连接组成。使用 Mininet[20]网络环境部署上述架构并执行实验。实验所用的网络拓扑结构由 Waxman[21]模型随机生成，使用参数 $\alpha = 0.2$，$\beta = 0.15$，拓扑中包含的节点总数为 280，该模型通过给定参数计算一个概率值来决定 2 个节点之间是否有直接相连的链路，所生成的拓扑结构具有一定的随机性。运行 Mininet 的宿主机硬件配置为 Inter i7 8700 CPU，32 GB；使用 Nvidia 1080ti GPU、TensorFlow 2.0[22]训练 DDPG 模型。

在上述网络中运行 4 种应用层服务协议，分别是 FTP（基于 TCP）、HTTP（基于 TCP）、WebRTC（基于 UDP）和 RTSP（选择基于 UDP），分别对应 4 种不同的服务场景，如表 3 所示。6 个 FTP 服务器和 3 个不同的 HTTP 服务器位于不同的主机上。一个 Web RTC 服务器和一个 RTSP 服务器也位于各自的主机之上。网络中可能的数据流数量为 2 196 个，均为客户主机与服务器间数据流，默认服务器间无通信。运行 FTP、HTTP、WebRTC 和 RTSP 这 4 项服务的主机配置均为 CentOS 7 系统、16 core 2.8 GHz、64 GB 内存，其余主机配置均为 Ubuntu 14.04 系统、8 core 2.2 GHz、8 GB 内存。所使用的 IP 地址空间为一个 B 类 IP 地址池，全部服务器和主机均随机分配一个固定 IP 地址。

表 3 服务场景与应用层协议

服务场景	应用层协议
文件传输	FTP
网页浏览	HTTP
视频通话	WebRTC
网络直播	RTSP

表 4 总结了所提方法中涉及的超参数的默认取值，各个超参数的释义在本文已有解释，此处不做赘述。

表 4 超参数默认取值

超参数	默认取值
T_{slot}	8 s
th	8
TH	10
μ	1
γ	2
R_{BAD}	−6
m	3
η	0.05
τ	0.01
f_c	10

与所提方法性能进行对比的 3 种方法为随机路由跳变（RRM）[8]、异常触发随机路由跳变（AT-RRM）[14]和时空随机优化路由跳变（SSO-RM）[12]。这 3 种方法均不使用 INT 技术收集网络状态信息，若需要网络状态信息，则在控制器的指导下收集所需的数据包以感知网络状态。

6.2 实验过程

在上述实验环境及设置下，分别对所提方法和 3 种对比方法执行安全性评价、QoS 评价、时间效率。在安全性评价实验中，以截获数据包占比和截获数据包离散度 2 项指标度量各方法的安全性；在 QoS 评价实验中，分别从时延表现和吞吐量表现 2 个方面评价各方法对网络整体合法 QoS 的保障水平。在时间效率评价实验中，由生成随机路由方案的时间开销评价各方法的时间效率，共执行 30 次实验，每次执行实验后初始化实验设置，取 30 次实验结果的平均值作为最终结果。

假设攻击方已经通过蠕虫传播或社会工程等手段将恶意窃听软件部署于网络中。本实验采用模拟方式执行窃听攻击，即按第 3 节所述威胁模型模拟窃听攻击行为。假设攻击方对网络中固定比例 $T_C = 0.3$ 的网络中的链路实施窃听，窃听位置可随机转移。当窃听攻击对某链路实施窃听，则经过该链路的数据包被截获率为 100%。忽略窃听行为对数据传输效率即设备复杂产生的影响。

在上述虚拟网络中执行所提防御方法，依据第 4 节描述的 DDPG 方法方案，实施 RRS 生成和更新；依据第 5 节描述数据包级粒度随机路由方案，在数

据平面实现数据包的随机转发，记录所有传输数据包的传输路径和转发时间点。将全部数据包传输记录与模拟运行的窃听攻击过程结合，即可计算实验结果评价所需的各项性能指标。

6.3　实验结果及评价

6.3.1　安全性

采用截获数据包占比（PPE, proportion of packet eavesdropped）和截获数据包离散度（IPD, intercepted packet dispersion）2 个性能指标衡量防御性能。PPE 即被截获的数据包数量占传输的数据包总量的比例。IPD 的定义如式(8)所示。

$$\mathrm{IPD}=\frac{\sum_{i=1}^{M-1}(\mathrm{index}_{i+1}-\mathrm{index}_i)}{M} \tag{8}$$

其中，M 表示某数据流中被截获的数据包数量，index_i 表示属于该数据流的第 i 个被截获的数据包在该流中的序号。

图 7 展示了所提方法和 3 种对比方法的 PPE 表现。从第 1 个防御周期到约第 20 个防御周期，DDPG 方法中的深层神经网络经训练至收敛。在收敛之后，所提方法的 PPE 表现优于其他 3 种方法。

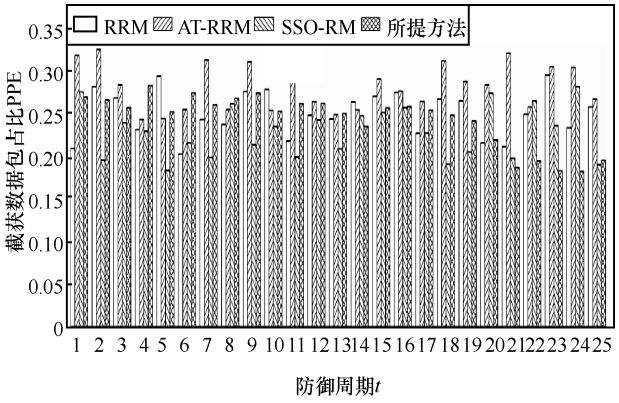


图 7　所提方法和 3 种对比方法的 PPE 表现

由于所提方法以接近数据包级粒度随机转发数据包，在任意时间点上数据流的路由路径存在“不确定性”。SSO-RM 方法在一个防御周期内仅有一个路由路径，路由路径相对确定。对于 AT-RRM 方法，由于窃听攻击不会引起明显的网络流量异常，因此路由跳变对攻击行为不敏感而难以做出及时的响应，导致 AT-RRM 的 PPE 表现甚至略低于经典的 RRM 方法。

所提方法与 3 种对比方法的 IPD 表现如表 5 所示。得益于更细的随机转发粒度，无论是对到达顺序敏感的 TCP，还是对到达顺序不敏感的 UDP，所提方法较 3 种对比方法均有更好的 IPD 表现。

表 5　所提方法与 3 种对比方法在不同协议下的 IPD 表现

方法	网络协议			
	TCP		UDP	
	FTP	HTTP	WebRTC	RTSP
RRM	3.190	1.797	3.711	3.854
AT-RRM	2.553	1.801	3.802	3.263
SSO-RM	3.217	1.911	3.786	3.299
所提方法	3.454	2.009	5.183	5.499

6.3.2　QoS

为了比较所提方法与 3 种对比方法在实施防御时对 QoS 的保障效果，实验比较整个网络吞吐量和时延表现。

1) 时延

网络整体的时延越低，表明数据包传输效率高，网络业务响应时间短。令 Del_t（单位为 ms）表示网络在第 t 个防御周期内的整体时延，其计算式如式(9)所示。

$$\mathrm{Del}_t=\frac{\mathrm{Td}_{\mathrm{dsp}}^t}{N_{\mathrm{dsp}}^t} \tag{9}$$

图 8 描绘了所提方法与 3 种对比方法在各个防御周期中的时延表现。自第 13 个防御周期起，所提方法较 3 种对比方法具有明显优势。自第 1 个防御周期至第 22 个防御周期，应用所提方法的网络整体时延呈下降趋势，说明该期间 DDPG 模型处于尚未收敛的学习阶段。

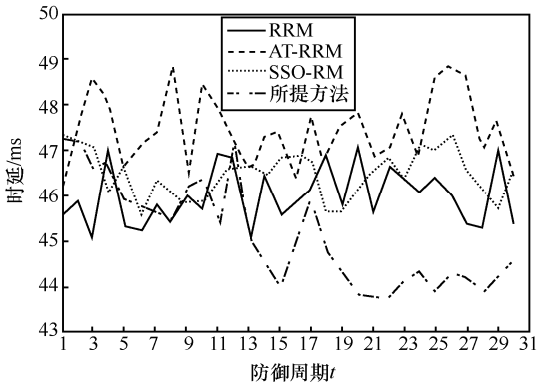


图 8　所提方法与 3 种对比方法在不同防御周期中的时延表现

2) 吞吐量

网络整体的吞吐量越高，表明带宽利用率越

高，对带宽需求高的服务中用户体验越好。令 Thr_t（单位为 Mbit/s）表示第 t 个防御周期中网络的整体吞吐量，其计算式如式(10)所示。

$$\mathrm{Thr}_t = \frac{\mathrm{Vol}_t}{T_{\mathrm{slot}}} \tag{10}$$

图 9 描绘了所提方法与 3 种对比方法在各个防御周期中的吞吐量表现。自第 8 个防御周期起，所提方法较 3 种对比方法具有明显优势。自第 1 个防御周期至第 23 个防御周期，应用所提方法的网络整体吞吐量呈上升趋势，说明该期间 DDPG 模型处于尚未收敛的学习阶段。

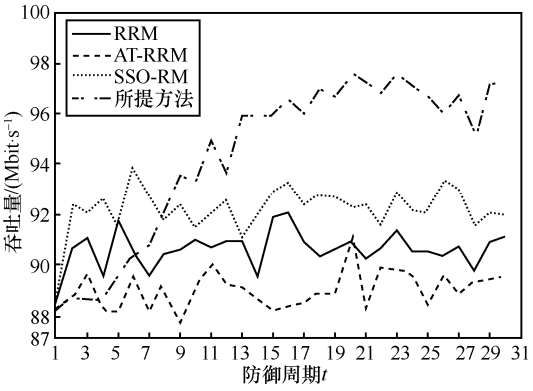


图 9 所提方法与 3 种对比方法在不同防御周期中的吞吐量表现

3) 数据包失序

在评估数据包失序问题的实验中，为衡量不同方法中数据包失序问题的严重程度，以触发 3-dup-ack 的数据包占比（PPR, proportion of packet reordering）为性能指标，占比越高表明数据包失序问题越严重。

在实验拓扑中随机选择 2 个主机，它们之间共有 6 条可行路由路径，该 6 条路由路径共有 44 条链路。在 44 条链路中，3 条链路由 2 条路由路径共享，一条链路由 3 条路由路径共享。所选主机之一充当 FTP、HTTP、WebRTC 或 RTSP 服务器。在 FTP 测试中，服务器向主机传输一个 1 GB 的文件。在测试 HTTP 中，用户主机不断访问服务器。在 WebRTC 和 RTSP 测试中，服务器分别执行视频通话和网络直播业务。

表 6 描述了所提方法和 3 种对比方法分别在 FTP、HTTP、WebRTC 和 RTSP 服务中触发 3-dup-ack 的数据包占比。可见，3 种对比方法中，数据包产生拥塞信号的比例为 0.075 9~0.141 3。所提方法在该指标上的表现仅为 0.003 9~0.005 1，所提方法明显优于 3 种对比方法。

6.3.3 时间效率

表 7 展示了所提方法与 3 种对比方法生成路由随机化方案的时间效率。可见，所提方法时间效率为 0.032 786 s，优于 3 种对比方法。

表 6 所提方法与 3 种对比方法的数据包失序评价

方法	FTP	HTTP	WebRTC	RTSP
RRM	0.126 3	0.122 2	0.141 3	0.135 2
AT-RRM	0.077 6	0.075 9	0.081 9	0.080 1
SSO-RM	0.115 4	0.119 9	0.113 8	0.120 1
所提方法	0.003 9	0.005 1	0.004 7	0.004 9

表 7 所提方法与 3 种对比方法的时间效率评价

方法	时间效率/s
RRM	0.051 499
AT-RRM	0.120 358
SSO-RM	0.061 295
所提方法	0.032 786

7 结束语

本文提出了一种基于 DDPG 方法的随机路由防御方法，从更细的路由随机粒度、更实时准确的网络状态感知和更强大的决策 3 个方面入手，提高了网络系统对抗窃听攻击的安全性，同时兼顾了对网络 QoS 的保障。

后续工作包括：首先，在 DDPG 模型的“状态”中添加有关实时攻击行为的信息，使所提方法能够进一步适应攻击者行为；其次，进一步优化 DDPG 方法，以提高方法的性能表现；最后，相关厂商合作探索将所提技术部署于大规模真实网络环境中，对所提方法的各方面表现进一步验证，不断提高实验结果的可靠性。

参考文献：

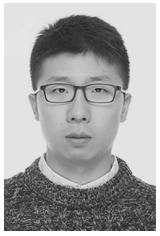
[1] 赛门铁克. 2018 年安全威胁趋势预测[J]. 网络安全和信息化, 2018(1): 103-105.
SYMANTEC. Security threat trend forecast in 2018[J]. Security & Informatization, 2018(1): 103-105.

[2] JAJODIA S, GHOSH A K, SWARUP V, et al. Moving target defense[M]. New York: Springer, 2011.

[3] YANG W, ZHENG Z Q, CHEN G R, et al. Security analysis of a distributed networked system under eavesdropping attacks[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2020, 67(7): 1254-1258.

[4] GURUNG S, CHAUHAN S. A survey of black-hole attack mitigation techniques in MANET: merits, drawbacks, and suitability[J]. Wireless Networks, 2020, 26(3): 1981-2011.
[5] SINGH M P, BHANDARI A. New-flow based DDoS attacks in SDN: Taxonomy, rationales, and research challenges[J]. Computer Communications, 2020, 154: 509-527.
[6] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. arXiv Preprint, arXiv:1509.02971, 2015.
[7] BOSSHART P, DALY D, GIBB G, et al. P4[J]. ACM SIGCOMM Computer Communication Review, 2014, 44(3): 87-95.
[8] DUAN Q, AL-SHAER E, JAFARIAN H. Efficient random route mutation considering flow and network constraints[C]//2013 IEEE Conference on Communications and Network Security. Piscataway: IEEE Press, 2013: 260-268.
[9] JAFARIAN J H, AL-SHAER E, DUAN Q. Formal approach for route agility against persistent attackers[C]//European Symposium on Research in Computer Security. Berlin: Springer, 2013: 237-254.
[10] ZHAO Z, GONG D F, LU B, et al. SDN-based double hopping communication against sniffer attack[J]. Mathematical Problems in Engineering, 2016, 2016: 1-13.
[11] ASEERI A, NETJINDA N, HEWETT R. Alleviating eavesdropping attacks in software-defined networking data plane[C]//Proceedings of the 12th Annual Conference on Cyber and Information Security Research. New York: ACM Press, 2017: 1-8.
[12] ZHOU Z, XU C Q, KUANG X H, et al. An efficient and agile spatio-temporal route mutation moving target defense mechanism[C]//2019 IEEE International Conference on Communications. Piscataway: IEEE Press, 2019: 1-6.
[13] DUAN Q, AL-SHAER E, CHATTERJEE S, et al. Proactive routing mutation against stealthy distributed denial of service attacks: metrics, modeling, and analysis[J]. The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 2018, 15(2): 219-230.
[14] LIU J, ZHANG H Q, GUO Z C. A defense mechanism of random routing mutation in SDN[J]. IEICE Transactions on Information and Systems, 2017, 100(5): 1046-1054.
[15] 雷程, 马多贺, 张红旗, 等. 基于最优路径跳变的网络移动目标防御技术[J]. 通信学报, 2017, 38(3): 133-143.
LEI C, MA D H, ZHANG H Q, et al. Network moving target defense technique based on optimal forwarding path migration[J]. Journal on Communications, 2017, 38(3): 133-143.
[16] ZHANG T, KUANG X H, ZHOU Z, et al. An intelligent route mutation mechanism against mixed attack based on security awareness[C]//2019 IEEE Global Communications Conference. Piscataway: IEEE Press, 2019: 1-6.
[17] ZHANG T, XU C Q, ZHANG B C, et al. DQ-RM: deep reinforcement learning-based route mutation scheme for multimedia services[C]//2020 International Wireless Communications and Mobile Computing. Piscataway: IEEE Press, 2020: 291-296.
[18] KIM C, SIVARAMAN A, KATTA N, et al. In-band network telemetry via programmable dataplanes[J]. ACM SIGCOMM, 2015, 17: 1-2.
[19] KANDULA S, KATABI D, SINHA S, et al. Dynamic load balancing without packet reordering[J]. ACM SIGCOMM Computer Communication Review, 2007, 37(2): 51-62.
[20] KAUR K, SINGH J, GHUMMAN N S. Mininet as software defined networking testing platform[C]//International Conference on Communiction, Computing & Systems. Piscataway: IEEE Press, 2014: 1-6.
[21] WAXMAN B M. Routing of multipoint connections[J]. IEEE Journal on Selected Areas in Communications, 1988, 6(9): 1617-1622.
[22] ABADI M, BARHAM P, CHEN J, et al. TensorFlow: a system for large-scale machine learning[C]//Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. Berkeley: USENIX Association, 2016: 265-283.

[作者简介]



徐潇雨（1992− ），男，江苏连云港人，信息工程大学博士生，主要研究方向为主动防御和智能决策。

胡浩（1989− ），男，安徽池州人，博士，信息工程大学讲师，主要研究方向为网络安全态势感知。

张红旗（1962− ），男，河北遵化人，博士，信息工程大学教授、博士生导师，主要研究方向为网络安全、风险评估、等级保护和信息安全管理等。

刘玉岭（1982− ），男，山东济阳人，博士，中国科学院信息工程研究所副教授，主要研究方向为网络安全测评和等级保护。