计算机科学
COMPUTER SCIENCE

针对网络流量测量的完整性干扰攻击与防御方法

郑海斌, 刘欣然, 陈晋音, 王鹏程, 王楦烨

引用本文

郑海斌, 刘欣然, 陈晋音, 王鹏程, 王楦烨. 针对网络流量测量的完整性干扰攻击与防御方法[J]. 计算机科学, 2024, 51(8): 420-428.

ZHENG Haibin, LIU Xinran, CHEN Jinyin, WANG Pengcheng, WANG Xuanye. Integrity Interference Attack and Defense Methods for Network Traffic Measurement [J]. Computer Science, 2024, 51(8): 420-428.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230500101

针对网络流量测量的完整性干扰攻击与防御方法

郑海斌[1,2] 刘欣然[1] 陈晋音[1,2] 王鹏程[1] 王楦烨[1]
1 浙江工业大学信息工程学院 杭州 310023
2 浙江工业大学网络空间安全研究院 杭州 310023
(haibinzheng320@gmail.com)

摘 要 近年来,网络测量在评估网络状态、提高网络自适应能力方面取得了较好的性能,被广泛运用于网络管理中。然而,目前的大规模网络中存在异常行为导致的网络流量数据污染问题。例如,自治系统中的恶意节点通过伪造恶意流量数据来故意操纵网络指标,影响网络测量,误导下游任务决策。基于此,首先提出完整性干扰攻击方法,通过修改流量矩阵的最小代价,利用多策略干扰生成器生成恶意扰动流量的攻击策略,实现干扰流量测量的目的。然后,通过一种混合对抗训练策略,设计在网络中抵御此类攻击的防御方法,实现流量测量模型的安全加固。实验中对攻击目标进行了相应的限定,验证了完整性干扰攻击在受限场景下的攻击有效性。并通过混合训练的方式进行对比实验,验证了常规模型的加固方法可以提升模型的鲁棒性。

关键词:网络流量测量;安全性;攻击可行性;攻击检测
中图分类号 TP391

Integrity Interference Attack and Defense Methods for Network Traffic Measurement

ZHENG Haibin[1,2],LIU Xinran[1],CHEN Jinyin[1,2],WANG Pengcheng[1] and WANG Xuanye[1]
1 College of Information Engineering,Zhejiang University of Technology,Hangzhou 310023,China
2 Institute of Cyberspace Security,Zhejiang University of Technology,Hangzhou 310023,China

Abstract In recent years,network measurement has achieved good performance in evaluating network status and improving network self-adaptability,and is widely used in network management. However,there is a problem of network traffic data pollution caused by abnormal behavior in the current large-scale network. For example,malicious nodes in autonomous systems intentionally manipulate network metrics by forging malicious traffic data,affecting network measurements and misleading downstream task decisions. Based on this,this paper first proposes an integrity jamming attack method. By modifying the minimum cost of the traffic matrix,a multi-strategy jamming generator is used to generate an attack strategy that maliciously disturbs traffic,so as to achieve the purpose of jamming traffic measurement. Then,by providing a hybrid adversarial training strategy,a defense method against such attacks in the network is designed to achieve security hardening of the traffic measurement model. In the experiment,the attack target is limited accordingly,and the effectiveness of the integrity interference attack in the restricted scenario is verified. And through the comparison of the mixed training method,the robustness of the reinforcement method of the conventional model is verified.

Keywords Network traffic measurement,Security,Attack feasibility,Attack detection

1 引言

网络流量测量是网络管理的基础,已被广泛应用于互联网加密通信[1]、流量工程[2]、网络安全管理[3]等领域。网络流量测量通过收集和分析单位时间区间内的网络流量数据,预测并得出能够表征网络流量行为变化、掌握内部局域网络的特征状态。近年来,研究人员提出了生成对抗网络[4]、图卷积神经网络[5]和EMD聚类[6]的网络流量预测方法。

在大规模复杂场景的网络环境中,通过部署大量节点进行流量测量的方法受到测量框架和成本的限制。测量方法中所测即所得的评估方式缺乏对数据污染场景的关注。如图1所示,在自治系统中存在着内部威胁,恶意节点通过提取当前网络的关键链路,伪造恶意流量数据来操纵网络指标,影响网络测量,从而误导下游任务决策。更有甚者可以通过后门感染路由器,造成大面积的网络基础设备瘫痪。

到稿日期:2023-05-16 返修日期:2023-08-21
基金项目:浙江省自然科学基金(LDQ23F020001);国家自然科学基金(62072406)
This work was supported by the Natural Science Foundation of Zhejiang Province,China(LDQ23F020001) and National Natural Science Foundation of China(62072406)
通信作者:陈晋音(chenjinyin@zjut.edu.cn)

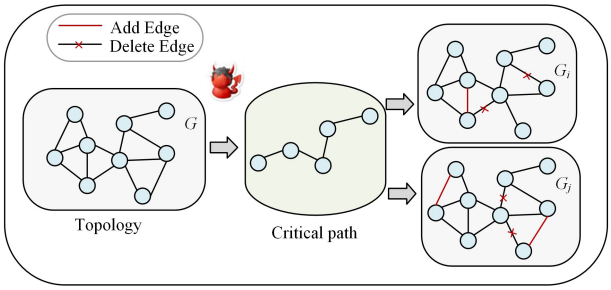


图 1 内部恶意节点的流量污染引发决策误导

Fig. 1 Traffic pollution from internal malicious nodes leads to misleading decisions

针对以上问题，提出了一种新的攻击策略，称为完整性扰动干扰。在利用网络流量测量中存在的所测即所得数据的漏洞前提下，验证现有网络流量测量方法存在的局限性与脆弱性。本文采用了 3 种针对不同目标的基本进攻策略。

1）选择性扰动策略（IA-selective），攻击者以一个或多个给定的受害者为目标，使得这些受害者成为共同造成网络流量测量问题的根本原因。

2）最大损害扰动策略（IA-max），攻击者在所有节点中找到最优的受害者集合，并对网络造成最大损害。

3）全局扰动策略（IA-global），通过网络特征表征，产生大量超出正常状态的数据包，从而迷惑网络拥有者，对网络造成最大损害。

本文首先分析了这些策略的可行性，并提出了检测和定位此类攻击的条件。通过网络数据集进行模拟攻击实验，以验证攻击的可能性与损害性。其次，基于实验观察与所提出的条件来检测和定位此类攻击。

本文的主要工作包括 4 个方面：

1）设计了一种黑盒的多策略干扰生成器，系统地构建了 3 种基本的攻击策略：选择性扰动策略、最大损害扰动策略、全局扰动策略。在保证隐蔽性的同时，影响下游任务，误导网络路由决策性能。

2）考虑受限条件下的完整性扰动干扰的有效性，在局部及全局特征下验证攻击的有效性。

3）从多方面考虑网络常用防御下完整性扰动干扰的有效性。

4）使用真实世界的数据集来评估各种设置的网络系统扰动对抗性攻击的威胁。实验结果表明，目前的网络流量测量容易受到外界攻击者的干扰。

2 相关工作

本章简要介绍网络流量测量方法和攻击方面的相关工作。

2.1 网络流量测量方法

目前已经提出了许多网络流量测量方法，主要可分为：直接测量法和估计推断法。

在传统网络中，网络管理员使用协议（SNMP，ICMP）或特殊诊断工具（Traceroute）周期性地查询单个网络组件，以获得网络指标（往返时延、丢包率与连通性等参数）。然而，由于缺乏对网络组件的支持功能、测量流量开销或在自治系统中被禁止，这种直接测量内部组件性能的方法并不总是可行的。

近年来，已有研究表明：网络测量的过程中可引入智能算法，通过在网络中设立特殊的监视节点（即监视器）来测量网络部分端点到端点间的流量，进而推断网络状态或预测网络变化趋势。例如，网络流量矩阵中包含网络中所有可能的起点终点矩阵预测（Origin And Destination，OD）路由器对之间的流量。由于对流量矩阵（Traffic Matrix，TM）进行直接测量是非常困难的，因此，研究者根据网络流量的自相似特性[7]，对 TM 进行估计或预测，而不是直接测量 TM。

基于网络流量建模是一种有效的测量网络流量的方法，Soule[8] 等和 Liu 等[9] 使用高斯分布、自回归综合移动平均（ARIMA）等统计和线性模型对网络流量进行建模，但传统模型不能很好地提取网络流量[7] 的非线性特性。Valadarsky 等[10]、Azzouni 等[11] 和 Zhao 等[12] 基于神经网络（Neural network，NN）建立 TM 序列和 OD 流序列的时间序列模型。他们通过比较预测的流量矩阵序列（Traffic Matrix Sequence，TMs）和实际值的预测误差来评价不同的预测方法，证明基于神经网络的方法比传统的线性模型具有更高的预测精度。预测误差小的预测方法并不总是适用于 TE，这些工作还没有研究不同预测方法对 TE 性能的影响。Liu 等[13] 通过神经网络进一步证实了预测各个 OD 流可以进一步提高预测精度，并通过不同 TE 场景下的预测性能以及预测时间来评估不同的预测方法。

2.2 网络流量攻击方法

当前的网络流量场景中存在诸多网络攻击行为，这些攻击行为可以分为两类：网络流量操控和网络流量混淆。

网络流量操控是攻击者常用的渗透和评估目标网络的一种手段，可以在规避网络防御的同时，实现网络攻击的目的。例如，利用 Scapy 伪造、操纵、捕获和重放网络数据包；利用模糊技术操纵网络流量[14]。然而，由于其可扩展性，Scapy 只能对有效负载进行操作；模糊器不具备避开网络审查的能力。

网络流量混淆作为另一种常用的攻击方法，可以产生恶意流量，从而导致网络防御系统误分类。Dyer 等[15-16] 提出了两个可编程的网络流量混淆系统，Marionette 和 FTE，可以产生恶意流量，导致网络防御系统误分类。与随机化、模仿和隧道化等个人通信混淆技术相比，他们的可编程系统可以为用户提供最适合其使用情况的混淆方法。然而，该系统不支持更广泛的功能，例如流量隧道。

现有的网络流量攻击方法大多倾向于根据有限状态机来选择最优攻击操作，对网络协议有着更高的要求，还未达到普适性的效果。

3 预备知识

本章介绍了网络相关的问题定义和攻击目标模型的概念，为后文的方法部分做简要铺垫。

3.1 问题定义

本节介绍了网络测量的定义和一般的攻击问题。为方便起见，表 1 简要总结了一些必要符号的定义。

定义 1（网络拓扑） 网络由节点和链路组成，因此，通常可用有向图 $G=(N,L)$ 来表示，其中 N 表示网络中路由器

节点的个数，L 表示节点之间的链路集合。设邻接矩阵 $\boldsymbol{A}^{N\times N}$ 表示节点之间的连接关系，邻接矩阵 $\boldsymbol{A}^{N\times N}$ 只包含元素 0 和 1。当元素为 0 时，表示节点之间没有连接；当元素为 1 时，表示节点之间有连接。邻接矩阵 $\boldsymbol{X}$ 表示节点链路之间的流量大小，$\boldsymbol{X}$ 的每一项定义为 $x_{i,j}$，也就是说，每个链路 $l_j \in L$ 与一个描述为节点 i 到 j 的流量度量关系相关联。

定义 2（网络流量测量） 网络流量测量的目标是根据实测的历史网络流量信息预测未来的网络流量信息。将这个过程定义为：假设上述网络中的一组用户通过 G 沿着一组路径发送流量，并可以通过预测模型得到网络变化趋势。测量流量矩阵序列用 $\boldsymbol{X}_t(t\in[1,T])$ 表示，其中 T 为测量的总时数。将测量模型抽象为 $\boldsymbol{Y}=g(\boldsymbol{X})$ 的形式，其含义为通过一系列历史数据 $(\boldsymbol{X}_{t-k},\boldsymbol{X}_{t-k+1},\cdots,\boldsymbol{X}_{t-1})$ 来实现对未来时刻的网络态势 $\boldsymbol{Y}$ 的评估，其中 $k\in[1,t]$，$g(\cdot)$ 为需要训练的预测模型。

定义 3（网络流量操纵） 假设网络 $G=(N,L)$ 中，攻击者试图通过目标节点选择一些网络关键链路来操纵某些流量，从而构造出扰动网络 $\bar{G}=(N,\bar{L},\boldsymbol{M})$，其中 $\boldsymbol{M}_{ij}$ 表示 $\bar{L}_{ij}\in\bar{L}$ 的操纵的策略。那么，流量矩阵中的 OD 流 $\overline{x_{ij}}\in\bar{\boldsymbol{X}}$ 也会受链路的影响而变化。因此，被操纵的网络 $\hat{G}=(N,\hat{L})$ 中链路 $\bar{L}$ 中流量矩阵变化可以定义为：

$$\hat{\boldsymbol{x}}_{ij}=\boldsymbol{x}_{ij}+\boldsymbol{x}_{ij}\bar{\boldsymbol{x}}_{ij} \tag{1}$$

此外，攻击者只能控制被攻击链路的性能，被攻击的目标节点可以很好地隐藏。例如，在网络异常检测任务中，它们被错误分类的概率较大。

完整性扰动干扰由潜在的对抗性样本发起，与正常样本的交替极小。因此，为了保证攻击的隐蔽性，攻击者必须隐藏网络中的扰动节点。在针对网络异常分类任务中，由于大多数节点的向量与原始网络中的向量保持一致，即能够正确分类，而目标节点会因为向量的显著变化而无意识地误分类。更值得注意的是，操纵网络几乎与原来的网络相同。

3.2 攻击目标

Azzouni 等[11]的研究表明，基于神经网络的 TM 预测方法比自回归移动平均（ARMA）、时间序列预测算法（Holt-Winters）和主成分分析等传统方法具有更好的性能。因此，主要攻击目标采用基于神经网络测量的方法。

Troia[17]等和 Ramakrishnan 等[18]验证了循环神经网络（RNN）在 TM 预测中的性能，典型代表有长短记忆网络（LSTM）和门控循环单元（GRU）。此外，由于归一化 TM 是一个二维矩阵，可以将其视为一幅灰度图像，因此 TM 预测问题可以转化为灰度图像序列的时间序列预测问题。在此建立了两个新的基于卷积神经网络（CNN）的图像序列预测模型。

长期循环卷积网络（Long-term Recurrent Convolutional Networks，LRCN）是一种新的图像序列预测模型，用于生成图像和视频的文本描述[19]。LRCN 使用 CNN 层进行图像特征提取，并基于 LSTM 进行序列预测。该方法对 TM 序列进行最小-最大归一化处理，并将归一化后的 TM 序列作为灰度图像处理。然后使用 LRCN 对图像序列进行建模，构建了包含 3 个 CNN 层和一个包含 100 个隐藏节点的 LSTM 层的 LRCN 模型。

时间卷积网络（Temporal Convolutional Network，TCN）是 CNN 架构的一项创新，它使用扩展的因果卷积和残留连接来显示更长的记忆，并且在几个序列建模任务[20]中比典型的循环架构（如 LSTM 和 GRU）表现得更好。该方法还将归一化的 TMs 视为灰度图像，并使用 TCN 对图像序列进行建模。

4 最佳攻击策略

本文提出的完整性扰动干扰通过结合不同的修改策略、攻击规模和隐蔽性约束，对多个网络测量任务进行攻击。图 2 展示了完整性扰动干扰对各种测量任务的处理过程，该系统由多策略干扰生成器（Multi-strategy Interference Generator，MIG）组成。完整性干扰通过对 MIG 进行训练，在给定隐蔽性约束下操纵链路流量属性，从而有效地攻击相应的目标模型。

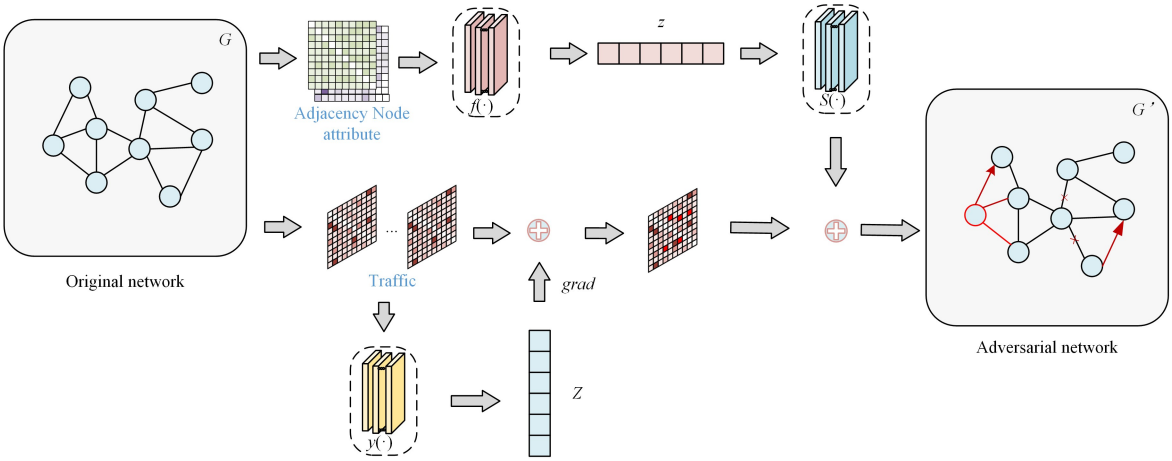


图 2 完整性扰动干扰工作流程

Fig. 2 Workflow of integrity perturbations interference

4.1 多策略干扰生成器

由于网络系统状态的确定性、静态性和同构性，现有的网络流量攻击方法倾向于根据有限状态机来选择最优攻击操作，而这些有限状态机通常是由特定的网络协议设计的。

这些方法的有限状态机对对抗实例的生成起着决定性的作用，对网络协议的要求可能更高。当应用于其他网络协议任务时，这些生成过程不能直接应用于其他函数，因为它们不是以可转移性为标准设计的。

考虑到攻击目标的多样性，多策略干扰生成器是一个值得研究的方向。利用多策略干扰生成器，即便在不知道目标测量模型结构的情况下，也可以获得良好的攻击效果。MIG为不同的网络测量任务快速生成对抗示例。接下来将介绍MIG的结构、多种攻击策略和多种隐蔽性约束。

4.2 生成器结构

MIG包含两个模块：一个特征提取器和一个网络生成器。该算法通过多种攻击策略和隐蔽性约束生成多种对抗实例。

为了更好地了解网络拓扑结构和流量特征，本文使用图卷积网络(GCN)和GRU作为网络特征提取器，并利用GCN对动态网络拓扑结构进行建模作为空间特征提取器，形式上，假设一个网络中有 N 个节点具有 m 维特征(或属性)。拓扑结构和节点属性可以分别用邻接矩阵 $\boldsymbol{A}\in\boldsymbol{R}^{N\times N}$ 与特征矩阵 $\boldsymbol{Z}\in\boldsymbol{R}^{N\times M}$ 表示。因此，GCN模型简单定义为：

$$\boldsymbol{H}^{(l+1)}=GCN(H^{(l)},\boldsymbol{A})=\sigma(\hat{\boldsymbol{A}}H^{(l)}\boldsymbol{W}^{(l)}) \tag{2}$$

其中，$\hat{\boldsymbol{A}}=\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}$，$\boldsymbol{A}$ 是邻接矩阵，$\tilde{\boldsymbol{A}}=\boldsymbol{A}+\mathbf{I}_N$ 是具有自连接拓扑关系的网络 G 的邻接矩阵，$\mathbf{I}_N$ 代表单位矩阵；$\tilde{\boldsymbol{D}}_{ii}=\sum_{j=1}^{N}\tilde{\boldsymbol{A}}$ 是拓扑邻接矩阵 $\tilde{\boldsymbol{A}}$ 的度矩阵；$\boldsymbol{W}^{(l)}$ 表示神经网络中的可训练参数，即权重矩阵。H 是每一层的特征，对于输入层的话，$H^{(0)}=\boldsymbol{Y}$，σ 是非线性激活函数Relu。考虑到不同网络状态下的链接关系，GCN模型对于每一个链接矩阵 $\boldsymbol{A}_t$，维护一个 $\boldsymbol{H}^{(l+1)}=GCN(H^{(l)},\boldsymbol{A})$。

将图结构和节点属性信息映射到 d 维节点表示形式 Z，定义为：

$$\boldsymbol{Z}=f^{\text{node}}(\boldsymbol{Y},\boldsymbol{A})=f(\hat{\boldsymbol{A}}\sigma(\hat{\boldsymbol{A}}\boldsymbol{Y}\boldsymbol{W}^{(0)})\boldsymbol{W}^{(1)}) \tag{3}$$

其中，$\boldsymbol{W}^{(0)}\in\boldsymbol{R}^{N\times H}$ 和 $\boldsymbol{W}^{(1)}\in\boldsymbol{R}^{H\times d}$ 表示隐层和输出层的可训练权矩阵，隐含层维度为 H。H 和 d 的值决定了学习的低维表示 $\boldsymbol{Z}$ 的质量。f 是softmax函数。为了掌握网络流量的时序特征关系，将原始流量数据 $\boldsymbol{X}_t(t\in[1,T])$ 输入门控制循环单元GRU。这里的更新门 u_t 和重置门 r_t 表达式如下：

$$u_t=\sigma(\boldsymbol{W}_u[f(\boldsymbol{A},\boldsymbol{X}_t),\boldsymbol{h}_{t-1}]+\boldsymbol{b}_u) \tag{4}$$

$$r_t=\sigma(\boldsymbol{W}_u[f(\boldsymbol{A},\boldsymbol{X}_t),\boldsymbol{h}_{t-1}]+\boldsymbol{b}_r) \tag{5}$$

其中，$f(\boldsymbol{A},\boldsymbol{X}_t)$ 表示图卷积过程，h_{t-1} 为隐藏层上一刻状态，$\boldsymbol{W}_u$，$\boldsymbol{b}_u$ 和 $\boldsymbol{b}_r$ 表示GRU的可训练参数。σ 是sigmoid函数，$\sigma(x)=1/(1+\mathrm{e}^{-x})$，$\tanh(\cdot)$ 为双曲正切函数。更新门 u_t 和重置门 $\boldsymbol{r}_t$ 共同决定隐藏层状态 h_t 的更新和重置。所述隐藏层状态 h_t 即为时序相关性特征：

$$\dot{\boldsymbol{y}}=\boldsymbol{h}_t=u_t*\boldsymbol{h}_{(t-1)}+(1-u_t)*\boldsymbol{c}_t \tag{6}$$

其中，$\boldsymbol{c}_t$ 为时间步 t 的候选隐藏层状态。

$$\boldsymbol{c}_t=\tanh(\boldsymbol{W}_u[f(\boldsymbol{A},\boldsymbol{X}_t),\boldsymbol{r}_t*\boldsymbol{h}_{(t-1)}]+\boldsymbol{b}_c) \tag{7}$$

其中，$\boldsymbol{W}_u$ 和 $\boldsymbol{b}_c$ 表示GRU的可训练参数，$\dot{\boldsymbol{y}}_i$ 表示GRU模型的输出。对于回归问题，使用欧氏距离(L2 loss)来评估所有训练样本。

$$L=\frac{1}{n}\sum_{i=1}^{n}(y_i-\dot{y}_i) \tag{8}$$

其中，n 表示网络中的节点数，y_i 表示网络流量的真实值，$\dot{y}_i$ 表示GRU模型的输出。在第 m 次迭代步骤中，对神经网络模型的权值 W_i 采用梯度下降的训练方式以更新模拟合方向。

$$\boldsymbol{W}_i^{m+1}=\boldsymbol{W}_i^{m}-\eta\frac{\partial L}{\partial W_i^m} \tag{9}$$

网络生成器：通过特征提取器获取网络空间特征 $\boldsymbol{Z}$ 和流量时序特征 $\boldsymbol{h}_t$ 后，使用一个展开矩阵 $\boldsymbol{W}_{ex}$ 将 $\boldsymbol{Z}$ 重构为连续对抗例子 $\boldsymbol{G}_c'$：

$$\boldsymbol{G}_c'=\begin{cases}\boldsymbol{A}_c=S((\boldsymbol{Z}\boldsymbol{W}_{ex}^{A}+(\boldsymbol{Z}\boldsymbol{W}_{ex}^{A})^{\mathrm{T}})/2)\\ \boldsymbol{X}_c'=S(\boldsymbol{Z}\boldsymbol{W}_{ex}^{X})\end{cases} \tag{10}$$

其中，$\boldsymbol{Z}\subset\boldsymbol{R}^{N\times d}$，$\boldsymbol{W}_{ex}^{A}\subset R^{N\times d}$，$\boldsymbol{W}_{ex}^{X}\subset R^{d\times D}$ 是网络结构 A 和节点属性 $\boldsymbol{X}$ 的展开矩阵。Sigmoid函数 S 将生成数据的元素值映射到[0，1]之间，然后得到网络结构 G'。

$$G'=\begin{cases}\boldsymbol{A}'=\mathrm{sign}(\boldsymbol{A}_c)\\ \boldsymbol{X}'=sign(\boldsymbol{X}_c')\end{cases} \tag{11}$$

其中，sign函数将大于0.5的值设置为1，其他值设置为0。

4.3 多策略攻击

在完整性干扰，设计了可选的修改策略和攻击规模，以满足不同攻击场景的需求。

根据式(9)，基于时序特征提取模型中的梯度信息更新权值矩阵，使模型不断优化，节点流量评估性能不断提高。由式(4)－式(6)可知，邻接矩阵 $\boldsymbol{X}_t(t\in[1,T])$ 是损失函数中的另一组变量。由式(2)和式(3)可知，网络拓扑提取的特征信息能够有效定位网络中的关键节点，这样就可以利用邻接矩阵的梯度信息生成精心准备的扰动，实现攻击，即导致流量评估偏离。

在训练好生成器中的时序特征提取模型的基础上，进一步设计了目标损耗函数 L_t。

$$L_t=\sum_{i=1}^{n}(y_{it}-\dot{y}_{it}(\boldsymbol{X}_t)) \tag{12}$$

该函数表示预测流量与目标链路 n_t 上流量的差值。该损失函数值越大，预测结果越差。然后计算目标损失函数 L_t 对网络中邻接矩阵元素 $\boldsymbol{x}_{ij}$ 的偏导数，进而得到所有的流量梯度矩阵 $\boldsymbol{g}$，表示为：

$$\boldsymbol{g}_{ij}=\frac{\partial L_t}{\partial \boldsymbol{x}_{ij}} \tag{13}$$

其中，$\boldsymbol{x}_{ij}$ 是网络流量矩阵 $\boldsymbol{X}$ 的元素。

本文的目标是最大化目标损失函数 L_t。沿着梯度方向的链接变化会使目标损失函 L_t 局部增长最快，使得目标网络的评估模型在精度上偏离真实值，影响下游任务。考虑到无向网络的邻接矩阵是对称的，对 $\boldsymbol{g}$ 进行对称化得到 $\hat{\boldsymbol{g}}$。

$$\hat{\boldsymbol{g}}_{ij}=\hat{\boldsymbol{g}}_{ji}=\begin{cases}\dfrac{\boldsymbol{g}_{ji}+\boldsymbol{g}_{ij}}{2}, & i\neq j\\ 0, & i=j\end{cases} \tag{14}$$

其中，$\hat{\boldsymbol{g}}$ 可以视为一种流量梯度网络(TGN)，其中每对节点可以表示流量梯度的正或负权值幅值关系。一个大于给定阈值 $\tilde{g}$ 的流量梯度 $\hat{\boldsymbol{g}}$ 表示对该链接(v_i,v_j)伪造流量数据注入来增加目标评估模型的损失函数，可表示如下：

$$\theta(\hat{g})=\begin{cases}1, & \hat{g}\geqslant\tilde{g}\\ 0, & \hat{g}<\tilde{g}\end{cases} \tag{15}$$

根据式(13),本文提出了一个扰动生成器模型来实现对原始网络的有效攻击。在该模型中,每次迭代中修改一个链接内的流量,该过程总共持续 K 次迭代。第 h 次迭代可以用以下步骤描述。

构建 TGN:根据式(13)和式(14)使用对抗性网络邻接矩阵$\hat{\boldsymbol{X}}^{h-1}$,生成第 $h-1$ 次迭代的流量梯度网络 $\hat{g}^{h-1}$,其中$\hat{\boldsymbol{X}}^{0}=\boldsymbol{X}$。

选择目标链路:根据 $\hat{g}^{h-1}$,选择涵盖在网络结构 G' 中最大绝对流量梯度的一对或多对节点对(v_i,v_j),需要注意的是,在给定阈值 $\tilde{g}$ 的情况下,大于或小于梯度,表示在原有网络中注入或不注入相应的虚假流量信息。

扰动实现:利用所选的节点对(v_i,v_j)干扰网络链路,生成扰动样本网络 $\hat{G}^{h}$,定义为:

$$\hat{x}_{ij}^{h}=\hat{x}_{ij}^{h-1}+\lambda\theta(\hat{g}_{ij}) \tag{16}$$

其中,$\hat{x}_{ij}^{h}$和 $\hat{x}_{ij}^{h-1}$分别表示$\hat{\boldsymbol{X}}^{h}$和$\hat{\boldsymbol{X}}^{h-1}$的元素,$\theta(\hat{g}_{ij})$表示所选的节点对$(v_i,v_j)$的流量梯度 $\hat{g}_{ij}$的映射值,λ 是干扰因子。

4.4 多隐蔽性约束

在实现高效和有效攻击的同时,还要确保在对抗例子中有难以察觉的扰动。一种选择是通过设置隐蔽性约束损失函数来限制生成对抗实例时的扰动。由于 MIG 是通过生成策略来生成对抗实例的,因此额外的隐蔽性约束的损失函数只能对 MIG 起到引导作用,即不能保证生成的对抗实例满足我们的隐蔽性要求。

本文在生成对抗算例后进行隐身约束的评估,这是一种独立于攻击过程的方式,有助于实现多样化的隐蔽性约束。在这里考虑以下隐蔽性约束。

攻击阈值 Δ:表示累加在流量矩阵上的变化幅度,并将其限制在阈值 Δ 以内,以约束对抗实例的总体扰动大小。

Hurst 指数:根据网络流量自相似描述可知,Hurst 指数可以用来衡量局部流量与整体流量的相似性,反映攻击前后流量自相似特性的变化,判断流量正常与否。在此基础上,其还限制了时序水平上的扰动。

这里,本文考虑使用 Hurst 指数迭代算法来要求:

$$\rho(k)=\frac{1}{2}((k+1)^{2H}-2k^{2H}+(k+1)^{2H}) \tag{17}$$

其中,k 表示采样间隔,H 表示 Hurst 指数值,$\rho(k)$为协方差和方差之商。

5 实验

本章将完整性干扰应用于不同的网络流量测量任务,并将攻击结果与其他基线进行比较。主要的研究问题如下:

RQ1:考虑不同的网络流量数据集,完整性扰动干扰攻击性能与现有的攻击方法相比如何?

RQ2:考虑不同的网络测量任务,完整性干扰的攻击性能与现有的攻击方法相比如何?

RQ3:考虑在不同受限条件下对完整性扰动干扰的攻击性能有什么影响?

RQ4:考虑现有的可以使用的防御方法对攻击性能有什么影响?

RQ5:完整性干扰攻击下对网络下游任务有什么影响?

5.1 实验设置

在不同网络测量攻击实验中引入的分割比,将每个数据集随机分成训练集、验证集和测试集。一旦生成的流量操纵能够成功地误导目标网络的网络测量任务,就认为攻击是成功的。此时,停止攻击过程并输出对抗示例。本文对 MIG 进行了不同干扰因子的攻击实验,并使用 Adam 优化器来优化生成器,学习率为[0.01,0.001]。其他关键超参数通过特征提取器的隐藏层维度中的超参数搜索来设置,$H\in\{16,32,64,128,328\}$。

在实验中,当 MIG 的干扰因子为$\lambda\in\{10\%,30\%,50\%\}$,才能使得生成的对抗性示例既能满足攻击需求,又能满足隐蔽性约束。另外,为了避免关键的超参数对模型训练的影响,对下述实验内容统一设置指标:生成器最终学习率设置为 0.01,特征提取器的隐藏层维数均设为 128,并使用 Pytorch 来实现;攻击目标模型学习率为 0.01,隐藏层维数为 100,训练批次为 100,训练批量为 50,攻击因子设置为 10%。

此外,实验环境为 i7-7700K 3.5GHzx8(CPU),TITAN Xp 12GiB(GPU),16GBx4 内存(DDR4)和 Ubuntu 16.04(OS)。

5.2 数据集

为了评估本文方法的有效性,本文在不同网络系统的 3 个真实数据集和 1 个仿真数据集上进行了实验。如表 1 所列,其中 Abilene 数据集包含美国主干网中的 12 个节点之间的流量信息,连接着近 200 所美国的大学,同时也连接着世界上的其他国家。该数据集将这些地址上后 11 位数字进行隐匿来保护用户的安全。CERNET 网络是中国四大骨干网之一,CERNET 校园网要为各大高校提供网络流量服务。CERNET 网络数据集的主干节点包含华北地区的清华大学、西北地区的西安交通大学等八大地区的网络中心。GEANT 是研究泛欧洲的网络,它承载着来自连接大学和研究机构和欧洲国家研究和教育网络的研究流量,使用 ISIS 计算其域内路由。GEANT 记录了每个数据集连续 5 周的数据,然后在数据序列上建立滑动窗口,以进行连续的输入和训练目标模型。如文献[21]中所建议的,对于每 W 个连续数据,将前 $W-1$ 个数据作为模型的输入,最后一个将被视为输出。

表 1 不同网络系统的真实数据集介绍

Table 1 Introduction of datasets for different network systems

Datasets	Nodes	Links
Abilene	12	30
CERNET	14	32
GEANT	23	74

其中,在 GEANT 网络中,由于链路容量的范围有限,因此利用率最高的链路往往是容量较低的那条链路,平均链路利用率可以反映网络中所有链路的状态。我们从周一开始每隔 6h 统计一周网络的最大链路负载和平均值,结果如图 3 所示,从中可以清楚地看到工作时间的负载峰值和休息日负载的减少。

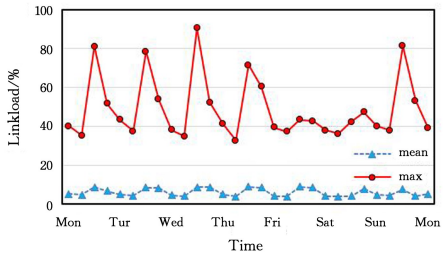


图 3 一周内链路负载的演变

Fig. 3 Evolution of link load in a week

5.3 对比算法

由于现有的网络流量攻击大多针对特定网络协议环境下行为，因此，本文将完整性干扰攻击方法与两种启发式的网络流量攻击方法进行了比较。在所研究的原始网络环境中，目标节点通过一系列链路与其他节点相连接。这些对比算法简述如下：

随机攻击(Random Attack)：随机操纵原有网络中若干链路中的流量统计特征，这是最简单的攻击方法。

模糊攻击(Fuzzy Attack)：设定流量模糊区间$[a,b]$，对于在模糊区间内的链路进行流量操纵以影响流量特征。

为了验证本文的完整性干扰方法的攻击能力，除了典型的循环架构的 LSTM[22] 和 GRU[23] 外，本文也比较了 3 种基线方法对 LRCN[24]，TCN[25] 等其他网络流量评估方法的攻击能力。

5.4 评价指标

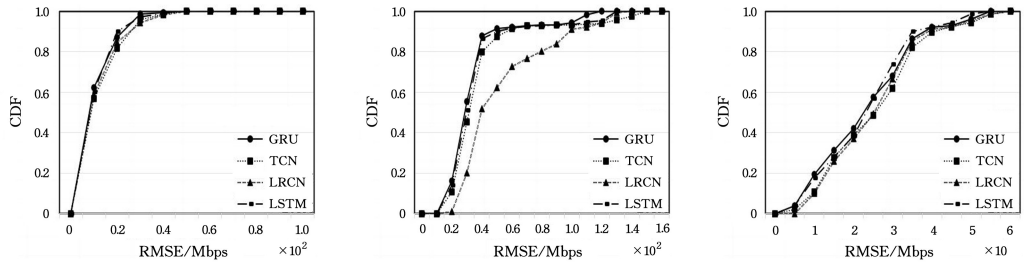
使用均方根误差(Root Mean Square Error，RMSE)(又称标准误差)来评估攻击性能，均方根误差主要反映预测值与实际值的均方根值的偏差。理想状态下，该值越小越好。预测流量矩阵 $\boldsymbol{X}'$ 与实际流量矩阵 $\boldsymbol{X}$ 之间的 RMSE 可以表示为：

$$RMSE=\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(x_{ij}-x'_{ij})^2} \tag{18}$$

其中，N 表示网络的节点号。

5.5 非受限场景下网络流量攻击

攻击目标节点为全局网络的若干个内部节点。下面介绍在不同数据集上，不同目标模型、不同攻击方法对测量性能变化的影响。首先将每个数据集切分为 3 部分：80%作为训练集，10%作为验证集，剩下的 10%作为测试集，训练不同流量测量神经网络模型作为攻击目标。其次，对比不同的约束条件下，完整性干扰方法和基线攻击方法的攻击效果之间的差异。最后绘制累积分布函数曲线，不同的神经网络模型的 RMSE 结果如图 4 所示。在 3 个数据集中，基于 RNN 的方法都能获得比 LRCN 和 TCN 更好的预测精度。

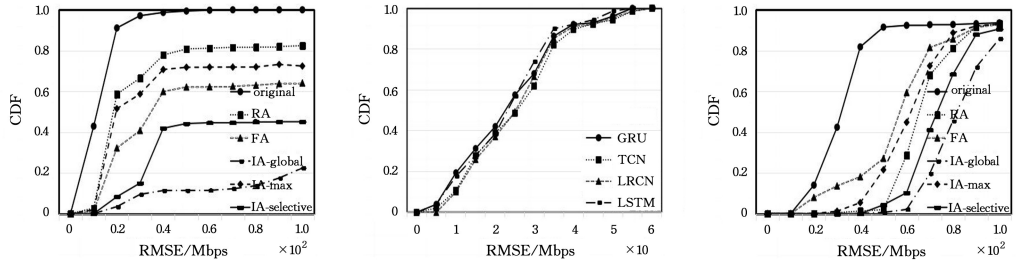


(a) Abilene (b) CERNET (c) GEANT

图 4 不同神经网络模型在未攻击场景下的 RMSE 结果

Fig. 4 RMSE of different neural network models in non-attack scenarios

对于问题 RQ1，如图 5 所示，在相同的约束条件下、相同的 LSTM 模型，在 3 个数据集中可以发现完整性干扰(IA)在大多数情况下比 RA 和 FA 的攻击效果更好。推测原因可能是，在流量攻击中，通过扰动干扰会影响链路本身流量特征，同时也会影响其周围原有链路的内联关系，使其具有更好的攻击效果。就干扰策略而言，对于 GEANT 数据集，IA-selective 攻击的有效性要稍大于 IA-global，究其原因，是其网络结构比较密集且攻击链路的流量梯度特征较为明显所致。

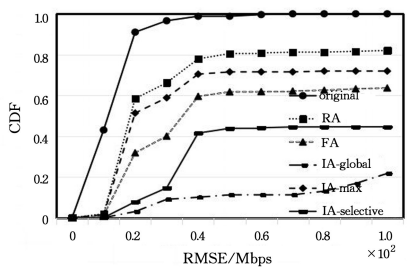


(a) Abilene (b) CERNET (c) GEANT

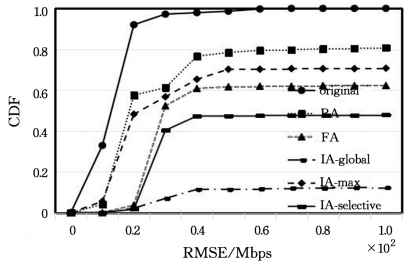
图 5 相同 LSTM 模型在不同数据集下的 RMSE 结果

Fig. 5 RMSE of the same LSTM model on different datasets

对于问题 RQ2，如图 6 和图 7 所示，在相同的约束条件下，相同的 Abilene 数据集，在 RNN 与 LRCN 和 TCN 模型的对比中，仍然可以发现完整性干扰(IA)在大多数情况下比 RA 和 FA 的攻击效果更好。对于 LRCN 模型，不同攻击方法的攻击有效性有所下降，这是由于 LRCN 对 TM 序列进行归一化处理，生成的灰度图像使得全局特征方差和减小，从而能有效去除恶意的操纵流量。尽管如此，完整性干扰(IA)在攻击任何可考虑的网络测量方法时仍然表现最佳。



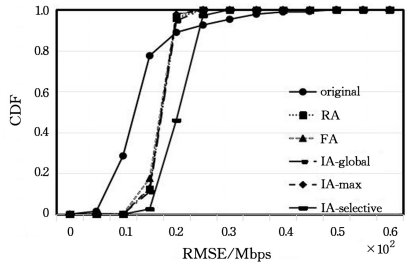
(a)LSTM



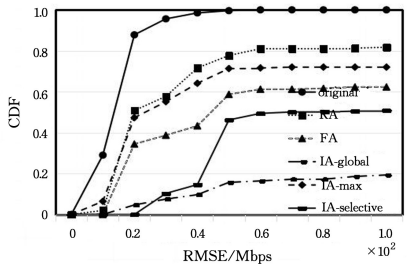
(b)GRU

图 6 Abilene 数据集下 RNN 模型的 RMSE 结果

Fig. 6 RMSE of RNN model on Abilene dataset



(a)LRCN



(b)TCN

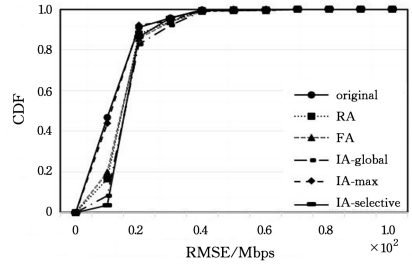
图 7 Abilene 数据集下 LRCN 和 TCN 模型的 RMSE 结果

Fig. 7 RMSE of LRCN and TCN models on Abilene dataset

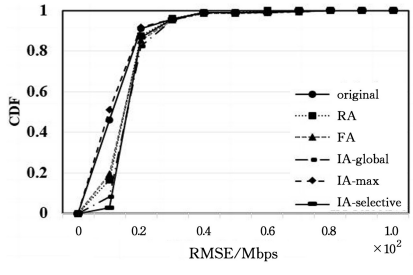
5.6 受限场景下网络流量攻击

考虑到实际网络场景中内部恶意节点渗透并控制的局限，以两个攻击节点为例，对于问题 RQ3，将以 Abilene 数据集进行分析。实验设置在原有的基础上，对攻击方法的攻击目标进行了相应的限定。

图 8 与图 9 展示了不同策略的生成器受限场景中生成的对抗样本在 4 种模型上的 RMSE 结果。可以看出，与全局攻击(IA-global)过程相比，选择性攻击(IA-selective)的攻击效果明显更佳。这可能是因为 IA-selective 攻击针对一个或多个特定的受害者进行攻击，实现了攻击效果的最大化，这与我们设定的攻击目标限定条件相符。



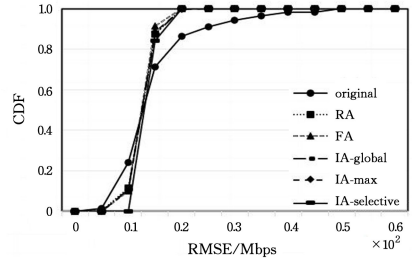
(a)LSTM



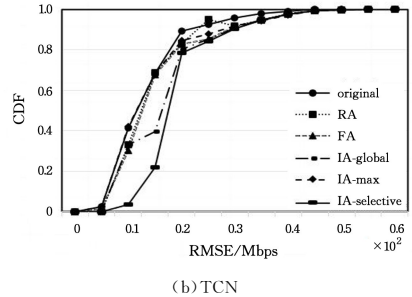
(b)GRU

图 8 Abilene 数据集受限场景下 RNN 模型的 RMSE 结果

Fig. 8 RMSE of RNN model in limited scenarios on Abilene dataset



(a)LRCN



(b)TCN

图 9 Abilene 数据集受限场景下 LRCN 与 TCN 模型的 RMSE 结果

Fig. 9 RMSE of LRCN and TCN models in limited scenarios on Abilene dataset

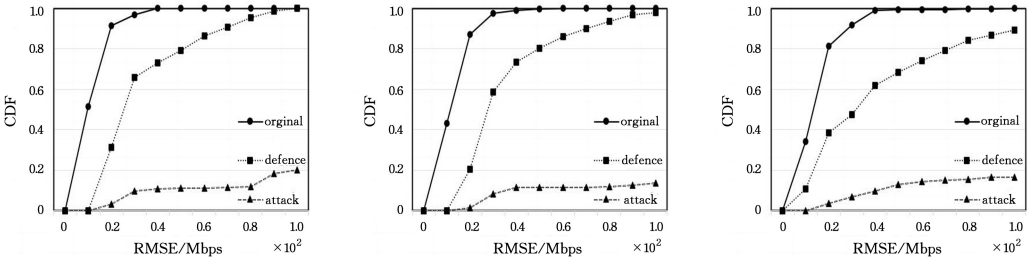
5.7 可能的防御

网络恶意攻击检测指通过一系列的技术和手段来识别和防御针对计算机网络系统的未授权攻击行为。随着互联网的普及和信息技术的快速发展，网络恶意攻击已经成为信息安全的重要威胁之一。网络恶意攻击检测技术可以帮助保护个人、企业和国家的网络安全，防止数据泄露、系统破坏和其他形式的损害。针对网络异常流量攻击，Chang 等[26]提出并实现了一种使用 Linux 内核观测技术 eBPF(Extended Berkeley Packet Filter) 与深度学习技术结合的基于网络流量特征分析的网络异常流量检测系统。Li 等[27]设计基于判别条件

变分自编码器与密度峰值聚类算法的入侵检测模型，具有更强的未知攻击检测能力，面对当前复杂网络环境，有更优的入侵检测性能。Shen 等[28]针对目前边缘网络中 DDoS 攻击检测方法性能不佳、未对卸载任务分类处理、对多属性的流量处理能力弱的问题，提出了一种基于任务分类的 Attention-1D-CNN DDoS 检测模型 TCA1C，对通信链路中的流量按不同的卸载任务进行分类，使单个任务受到攻击时不会影响整个链路中计算任务卸载的安全性，再对同一任务下的流量提取属性值并进行归一化处理。

对于问题 RQ4，即考虑到实际网络场景中存在防御方法的前提下，评估攻击的有效性。而常规的模型加固方法是通过对抗训练的方式来提升模型鲁棒性，因此，本文引入混合对抗训练的方式来进行对比。以 LSTM 模型为例，具体方案如下：在训练阶段，将随机生成器产生的扰动输入测量模型中进行时序数据的评估，经全连接层输出得到高维的时序特征。同时，生成的扰动样本将被输入鉴别器中，同样用测量模型进行高维时序特征提取。若经过全连接层的输出特征与原特征差异较大，则对生成器的参数进行更新迭代和优化。生成器和鉴别器通过两者协同对抗训练，使模型的安全性得到加固。

为了评估加固模型前后完整性干扰方法对模型攻击的有效性，以非受限场景下 IA-global 攻击为对象，以 Abilene 数据集为例进行分析。图 10 所给出了安全加固后的不同模型在 Abilene 数据集的对抗样本上的 RMSE 结果，可以发现，经防御加固后的模型在对抗样本上的攻击有效性有一定程度的减弱。



(a)LSTM (b)GRU (c)TCN

图 10 Abilene 数据集上非受限场景下加固模型前后的 RMSE 结果

Fig. 10 RMSE of models before and after reinforcement in limited scenarios on Abilene dataset

5.8 案例研究

考虑到真实网络场景中，网络攻击行为对网络下游任务的影响，对于问题 RQ5，首先，固定总流量大小为 9000 bps，网络拓扑结构使用 14 个节点和 21 个全双工链路的无标度网络结构，并使用重力模型初始化网络流量矩阵。具体网络拓扑结构如图 11 所示。在网络中，为了模拟真实的网络场景的需求，本文设计了一种融合路由状态及流量态势的 SDN 网络控制器，借助多维度网络状态的评估值实现路由决策任务。实验中路由决策器使用 OSPF 协议，通过定义链边的权重来决定某条链路与节点的流量大小从而模拟真实网络场景。

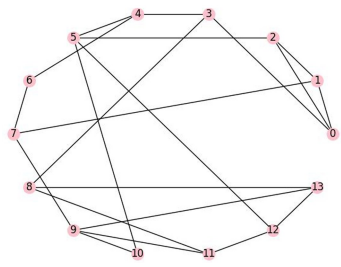


图 11 网络拓扑结构

Fig. 11 Network topology

以 LSTM 模型为例，表 2 列出了测量模型加固前后生成器对下游路由全局延时在不同攻击测量的影响，可以发现，针对未加固模型，随着干扰生成器攻击范围的扩大，下游路由决策得到的全局时延有所提高，表明了完整性干扰攻击下对网络下游任务的攻击有效性。对比加固模型场景下的情况，对下游路由的影响有所减小，但相比原始未干扰场景，还有一定差距。

表 2 LSTM 模型下测量模型加固前后生成器对下游路由全局延时在不同攻击测量的影响

Table 2 Influence of generator on global latency measurement of downstream routes before and after model reinforcement

Model	Perturbation method	Global average delay/s
Original LSTM	Original	0.286126
	IA-max	0.325147
	IA-selective	3.227173
	IA-global	4.479295
Enhanced LSTM	IA-selective	2.307154
	IA-global	3.514525

结束语 本文从理论和实验结果的角度来分析针对网络流量测量的完整性干扰攻击的可能性，并评价了网络流量测量攻击对下游任务及常用防御方法的影响，提出了减少这些攻击的方法。本文的研究结果表明，当前网络流量测量中缺乏对网络数据污染的关注，不应该简单地信任度量的网络数据指标，而应该始终意识到度量完整性攻击，并小心地重新审视各种应用程序中存在的安全性设计问题。未来可进一步开展以下研究：1)使用基于机器学习的异常检测算法对恶意流量进行检测，并且通过引入新的特征表示方法，如改进聚类算法、优化异常检测模型等来更加准确地检测物联网网络中的恶意流量；2)当前的恶意流量检测主要基于网络流量数据，但是恶意行为可能通过传感器数据、日志数据等多种流量数据进行。未来的研究可以探索如何将多模态数据进行融合，提高恶意流量检测的准确性和鲁棒性。

参 考 文 献

[1] PAPADOGIANNAKI E, IOANNIDIS S. A survey on encrypted network traffic analysis applications, techniques, and counter-

measures[J]. ACM Computing Surveys(CSUR), 2021, 54(6): 1-35.
[2] XIAO Y, LIU J, WU J, et al. Leveraging deep reinforcement learning for traffic engineering: A survey[J]. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2064-2097.
[3] ABBASI M, SHAHRAKI A, TAHERKORDI A. Deep learning for network traffic monitoring and analysis(NTMA): A survey [J]. Computer Communications, 2021, 170: 19-41.
[4] GAO Z Y, WANG T J, WANG Y, et al. Traffic Prediction Method for 5G Network Based on Generative Adversarial Network[J]. Computer Science, 2022, 49(4): 321-328.
[5] SONG Y L, LV G H, WANG G Z, et al. SDN Traffic Prediction Based on Graph Convolutional Network[J]. Computer Science, 2021, 48(S1): 392-397.
[6] YAO L S, LIU D, PEI Z F, et al. Real-time Network Traffic Prediction Model Based on EMD and Clustering[J]. Computer Science, 2020, 47(S2): 316-320.
[7] LI M, HAN D, YIN X, et al. Design and implementation of an anomaly network traffic detection model integrating temporal and spatial features[J]. Security and Communication Networks, 2021, 2021: 1-15.
[8] SOULE A, LAKHINA A, TAFT N, et al. Traffic matrices: balancing measurements, inference and modeling[C]// Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems. 2005: 362-373.
[9] LIU W, HONG A, OU L, et al. Prediction and correction of traffic matrix in an IP backbone network[C]// 2014 IEEE 33rd International Performance Computing and Communications Conference(IPCCC). IEEE, 2014: 1-9.
[10] VALADARSKY A, SCHAPIRA M, SHAHAF D, et al. Learning to route[C]// Proceedings of the 16th ACM Workshop on Hot Topics in Networks. 2017: 185-191.
[11] AZZOUNI A, PUJOLLE G. NeuTM: A neural network-based framework for traffic matrix prediction in SDN[C]// NOMS 2018—2018 IEEE/IFIP Network Operations and Management Symposium. IEEE, 2018: 1-5.
[12] ZHAO J, QU H, ZHAO J, et al. Towards traffic matrix prediction with LSTM recurrent neural networks[J]. Electronics Letters, 2018, 54(9): 566-568.
[13] LIU Z, WANG Z, YIN X, et al. Traffic matrix prediction based on deep learning for dynamic traffic engineering[C]// 2019 IEEE Symposium on Computers and Communications(ISCC). IEEE, 2019: 1-7.
[14] HANG L, KIM B H, KIM D H. A transaction traffic control approach based on fuzzy logic to improve hyperledger fabric performance[J]. Wireless Communications and Mobile Computing, 2022, 2022: 1-19.
[15] DYER K P, COULL S E, SHRIMPTON T. Marionette: A programmable network traffic obfuscation system[C]// 24th USENIX Security Symposium(USENIX Security 15). 2015: 367-382.
[16] DYER K P, COULL S E, RISTENPART T, et al. Protocol misidentification made easy with format-transforming encryption [C]// Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. 2013: 61-72.
[17] TROIA S, ALVIZU R, ZHOU Y, et al. Deep learning-based traffic prediction for network optimization[C]// 2018 20th International Conference on Transparent Optical Networks(ICTON). IEEE, 2018: 1-4.
[18] RAMAKRISHNAN N, SONI T. Network traffic prediction using recurrent neural networks[C]// 2018 17th IEEE International Conference on Machine Learning and Applications(ICMLA). IEEE, 2018: 187-193.
[19] DONAHUE J, ANNE HENDRICKS L, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2625-2634.
[20] LEA C, FLYNN M D, VIDAL R, et al. Temporal convolutional networks for action segmentation and detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 156-165.
[21] NIE L, JIANG D, GUO L, et al. Traffic matrix prediction and estimation based on deep learning in large-scale IP backbone networks[J]. Journal of Network and Computer Applications, 2016, 76: 16-22.
[22] BI J, ZHANG X, YUAN H, et al. A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM[J]. IEEE Transactions on Automation Science and Engineering, 2021, 19(3): 1869-1879.
[23] LI N, HU L, DENG Z L, et al. Research on GRU neural network Satellite traffic prediction based on transfer learning[J]. Wireless Personal Communications, 2021, 118: 815-827.
[24] RAI A, ALEEM A, GORE M M. Employing LRCN model for application classification in SDN[M]// Soft Computing for Problem Solving: Proceedings of SocProS 2020, Volume 2. Singapore: Springer Singapore, 2021: 347-359.
[25] LIU X, LIU Z A, ZHANG Y L, et al. TCN enhanced novel malicious traffic detection for IoT devices[J]. Connection Science, 2022, 34(1): 1322-1341.
[26] 昌武洋, 付雄, 王俊昌. 基于 eBPF 与 LSTM 的 DDoS 攻击检测系统[J]. 重庆工商大学学报(自然科学版), 2023, 40(2): 36-43.
[27] LI D H, GE L N, WANG Z, et al. Research on Network Intrusion Detection Model Combining DCVAE and DPC[J]. Journal of Chinese Computer Systems, 2024, 45(4): 998-1006.
[28] SHEN X Y, JI W F, LI Y Q, et al. TCA1C DDoS Detection Model for Edge Computing[J]. Computer Engineering, 2024, 50(1): 198-205.

ZHENG Haibin, born in 1995, Ph.D, lecturer. His main research interests include deep learning and artificial intelligence security.

CHEN Jinyin, born in 1982, Ph.D, professor. Her main research interests include artificial intelligence security, graph data mining and evolutionary computing.

(责任编辑:何杨)