针对 LTE-A 网络中的 DDoS 攻击流量检测模型

龚宇翔[1], 曹　进[1], 付玉龙[1], 郭　敏[2]

[1]西安电子科技大学 网络与信息安全学院 西安 中国 710126
[2]北京计算机技术及应用研究所 北京 中国 100854

摘要　近年来, 4G LTE-A 技术发展迅猛, 移动设备的普及以及各种承载于 4G 网络的业务和应用已经成为我们日常不可或缺的部分。但网络攻击技术也不断的在发展, 特别是近年来针对 4G LTE-A 网络的攻击技术的不断演进, 已成为危害人们切身利益的关键问题。DDoS 作为 DoS 攻击的一种, 对网络带来了更大的危害, 因此需要研究一种攻击检测模型。文章提出了一个针对 LTE-A 网络中的 DDoS 攻击流量检测模型, 模型利用熵作为特征之一, 并使用随机森林算法训练模型分类器, 可将其部署在 eNB 上对流经该 eNB 的 DDoS 流量进行识别。通过验证, 所提出的模型的检测准确率可达 99.956%。

关键词　机器学习; 随机森林; DDoS; LTE 网络; 熵
中图法分类号　TP393　DOI 号 10.19363/J.cnki.cn10-1380/tn.2019.01.03

A DDoS attack detection model for LTE-A network

GONG Yuxiang[1], CAO Jin[1], FU Yulong[1], GUO Min[2]

[1]School of Cyber Engineering, XiDian University, Xi'an 710126, China
[2]Beijing Computer Technology and Application Institute, Beijing 100854, China

Abstract　In recent years, 4G LTE-A technology has developed rapidly, and the popularity of mobile devices and various services based on 4G networks have become an indispensables part of our daily life. However, attack means is also constantly developing. The continuous evolution of attack means for 4G LTE-A networks in recent years has become a key issue that threatens our legal right. DDoS is a kind of denial of service attack, which brings more harm. Therefore, it is necessary to study an attack detection model. In this paper, a DDoS attack detection model for LTE-A network has been proposed. The model uses entropy as one of the features and uses random forest algorithm to train a classifier which can be equipped on an eNB to recognize the DDoS flow through the eNB. The experiment result shows that the detection accuracy of the proposed model can reach to 99.956%.

Key words　machine learning; random forest; DDoS; LTE network; entropy

1　简介

近年来, 4G LTE-A 技术发展迅猛, 移动设备如智能手机已经成为我们日常不可或缺的部分。相比于 3G 通信技术而言, 利用长期演进技术的 LTE-A 网络主要发展目标是将 LTE 发展成为 LTE-A 进而提高通信容量[1]。它为数十亿用户提供高级服务, 其带宽更高, 频谱效率更高, 延迟低于传统蜂窝网络。目前 5G 网络短期内依然会在 LTE-A 网络的基础上进行发展[2]。因此, 针对 LTE-A 网络的安全性研究是有必要的。由于它基于全 IP 异构架构, 因而容易遭受多种新型的攻击和威胁, 特别是 DDoS 攻击。DoS(Denial of Service)攻击即拒绝服务攻击主要消耗网络资源进而导致目标网络中的合法用户的通信时延显著提升甚至无法使用正常的网络服务[3-4]。这种攻击通过流式数据包传输, 从而降低了网络的处理能力, 导致网络拒绝合法用户的访问。DDoS(Distributed Denial of Service)攻击全称分布式拒绝服务攻击, 它是 DoS 攻击的一种, 带来的危害也更大。攻击者利用多个节点对同一目标进行 DoS 攻击, 可以产生大量的恶意数据包或不良请求, 更容易导致服务中断和目标网络崩溃[5]。通常来说, 区分合法数据包和攻击数据包是较为困难的, 因为 DDoS 的流量比较集中, 而没有一个显著于普通数据包的特征可以用于攻击的区分、检测和预防。有研究小组指出[6], 在 2017 上半

通讯作者：龚宇翔, 在读硕士, Email:gyx215@outlook.com。
本课题得到国家重点研发计划项目（No.2016YFB0800700）与国家自然科学基金项目（No.61772404，No.61602359）资助。
收稿日期: 2018-09-29; 修改日期: 2018-11-07; 定稿日期: 2018-12-06

年从各类攻击流量大小占比来看, SYN 泛洪攻击和 UDP 泛洪攻击依然是流量最大的两种攻击类型, SYN 泛洪攻击流量占比高达 56%, UDP 泛洪攻击流量占比为 23.3%。与 2016 年相比, SYN 泛洪攻击上升 7%, UDP 泛洪流量占比减少 6.3%。在大流量攻击方面, SYN 泛洪攻击明显增多。尤其在大于 300Gbps 的超大流量攻击中, SYN 泛洪攻击占比高达 91.3%, 相比去年增长了 52.3%。与此同时, UDP 泛洪攻击在大于 300Gbps 的超大流量攻击占比 8.7%, 相比去年下降 34.9%。由此可见, DDoS 泛洪攻击在网络各类攻击当中仍然占据很大比例, 并且仍有增加的趋势。近几年来, 新型的智能设备大量发展, 物联网(IoT)设备增长速度迅猛, 同时, 它们也是计算机网络用于 DDoS 攻击时流量放大的工具[7-31]。文献[32-33]表明, Ad-hoc 车辆自组织网易受 DDoS 攻击, 文章提出了检测 DDoS 攻击的模型以抵御 Ad-hoc 网络中的 DDoS 攻击。甚至包括 IEEE 802.15.4 低速无线个域网也会遭受 DDoS 攻击[34]。LTE-A 网络可以用于高速传输和 Ad-hoc 车辆自组织网络等场景中, 从上面可以看到, 各类无线网络均已出现 DDoS 攻击, 并且文献[13,35-37]提出, 目前的 LTE-A 网络在诸多方面遭受漏洞, 各层协议都有遭受攻击的可能, DDoS 攻击作为其中一种较为严重的攻击, 文献[13]指出, DDoS 也会在很大程度威胁到 LTE-A 网络的安全, 然而针对 LTE-A 网络 DDoS 攻击检测的模型或方法还较为欠缺。因此, 需要研究一种针对 LTE-A 网络中的攻击检测模型, 并利用该模型可以准确的检测 DDoS 攻击并对其进行防御。本文提出了一种针对 LTE-A 网络中的 DDoS 攻击流量检测模型, 利用部署在 eNB 上的基于随机森林的机器学习模型分类器, 使用时间窗口内的信息熵作为重要的特征之一, 可以有效检测 LTE-A 网络当中的 DDoS 流量。训练和测试数据集由运行在 Ubuntu 系统上的 NS-3 仿真生成。本文提出的模型可以有效检测出 LTE-A 网络中的 TCP SYN 泛洪攻击与 UDP 泛洪攻击。通过十折交叉验证, 准确率达 99.956%。

2 研究背景

在 LTE 网络中, 攻击者可以攻击接入网或核心网。在接入网方面, 主要攻击类型为信令攻击。L. Qiang 等人[9]提出了几种攻击, 它们会导致服务网关(S-GW)过载进而使正常用户无法建立承载。R. Bassil 等人[10]指出如下事实: 每个用户设备(UE)具有启动多达八个专用承载的能力加剧了信令的开销, 这使得 DoS 攻击成为可能。文献[11]指出如果一些相互配合的恶意用户反复地请求建立和终止承载, 网络资源会变得非常紧张并且网络服务质量(QoS)会下降。攻击者还可以利用僵尸网络同时下载或上传大量的数据或文件对某一小区的上传或下载链路进行攻击[12], 研究表明, 在上述攻击环境下, 只要某小区内的恶意程序传播 6%的 UE 用户即可对该小区实施攻击。在核心网方面, 攻击者可以利用僵尸网络对 LTE 无线接入网进行攻击, 攻击者通过伪造大量的虚假接入和分离信令, 并重复启动附着过程来泛洪移动管理实体(MME), 服务数据网关(S-GW)和分组数据网关(P-GW) [8]。文献[13]提出了攻击者可以利用发送大量的 SIP 信息给 P-CSCF 而耗尽 VoLTE 服务的资源。由于 LTE 网络是全 IP 网络, 攻击者也可以实施 TCP/SYN 泛洪攻击[13], 攻击者通过发送大量的 TCP/SYN 连接请求信息给目标服务器, 但是针对目标服务器的 SYN-ACK 信息并不回复 ACK 信息, 服务器便会等待用户。所以直到超时之前该服务器资源都会被占用。

很多研究者将目光放在机器学习上, 试图利用机器学习识别 DDoS 攻击流量。文献[14]利用 KNN, 支持向量机, 决策树和随机森林对 IoT 流量进行 DDoS 攻击检测识别, 准确率高达 99.9%。Carl L 等人[15]利用机器学习检测基于 IRC 的僵尸网络。作者比较了 J48 决策树, 朴素贝叶斯和贝叶斯网络分类器的检测效果。由于 IRC 僵尸网使用 TCP 协议, 于是作者在研究时只保留 TCP 数据包并且因为很多数据包载荷都在应用层加密, 于是忽略载荷只保留报头部分。这是十分具有借鉴意义的。文献[16]的工作利用了连接持续时间, TCP 数据包大小和 GET/POST 参数数量作为特征试图识别基于 Android 设备的僵尸网络, KNN 算法可以取得较高的准确率。

信息熵是一种衡量信息不确定性的统计方法。在网络流量中, 信息熵通过单一值度量来捕获流量的分布变化[21]。而对这些变化的充分观察可以清楚地揭示网络中的异常[22-24]已经被广泛证明。信息熵可以生成有用的流量分类功能, 并且在最近的 DDoS 攻击检测机制中得到了广泛的应用[17-20]。

3 基础知识

本章会简单介绍本文中需要用到的部分基础知识。

3.1 LTE-A 网络架构

LTE-A 网络基础架构如图 1 所示, 其中各个实体的功能如下:

UE: 用户设备, 一般为智能手机。

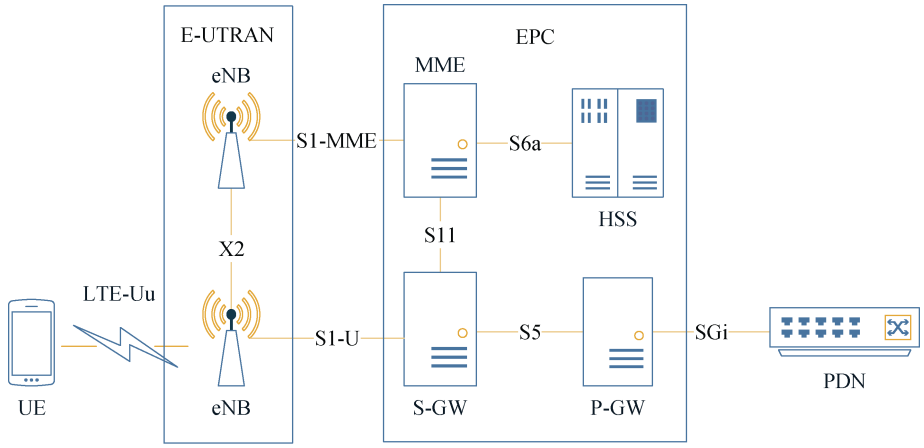

图 1　LTE 网络基础架构

Figure 1　LTE architecture

eNB: 演进型基站, 为 UE 提供无线接口并执行无线资源管理, 例如动态资源分配, 无线准入控制等。

MME: 移动性管理实体, 它是 E-UTRAN(演进通用移动通信系统陆地无线接入网)的主要控制实体, 它与 HSS 通信以进行用户认证和用户配置文件下载, 并使用 NAS 信令为 UE 提供 EPS 移动管理(EMM)和 EPS 会话管理(ESM)功能。主要负责空闲模式 UE 的定位, 寻呼和中继等工作。

HSS: 归属用户服务器, 它是存储用户配置文件的中央数据库, 主要向 MME 提供用户认证信息和配置文件。它存储 UE 的身份(例如, IMSI 和 IMEI)和订阅数据(例如, QoS 简档)及其从其生成认证质询的加密主密钥以及用于每个订户的对称会话密钥。

S-GW: 服务网关, 它是 E-UTRAN 接口的另一端, 主要是用于 eNB 间和 3GPP 网络间数据连接切换的本地移动锚点。

P-GW: 分组数据网关, 它通过从 PDN 的地址空间分配 IP 地址来向 UE 提供对 PDN 的访问。P-GW 用作 3GPP 和非 3GPP 网络之间的切换的移动锚点。

PDN: 分组数据网络, 一般为英特网。

3.2　TCP SYN 与 UDP DDoS 攻击

TCP 通信双方建立通信连接时需要进行三次握手, TCP SYN 的 DDoS 泛洪攻击正是利用了这个三次握手对某节点进行攻击。示意图如图 2 所示。在正常情况下, 通过 TCP 三次握手, 通信双方可以建立通信连接进行通信。而在 TCP SYN 泛洪攻击中, 攻击者通

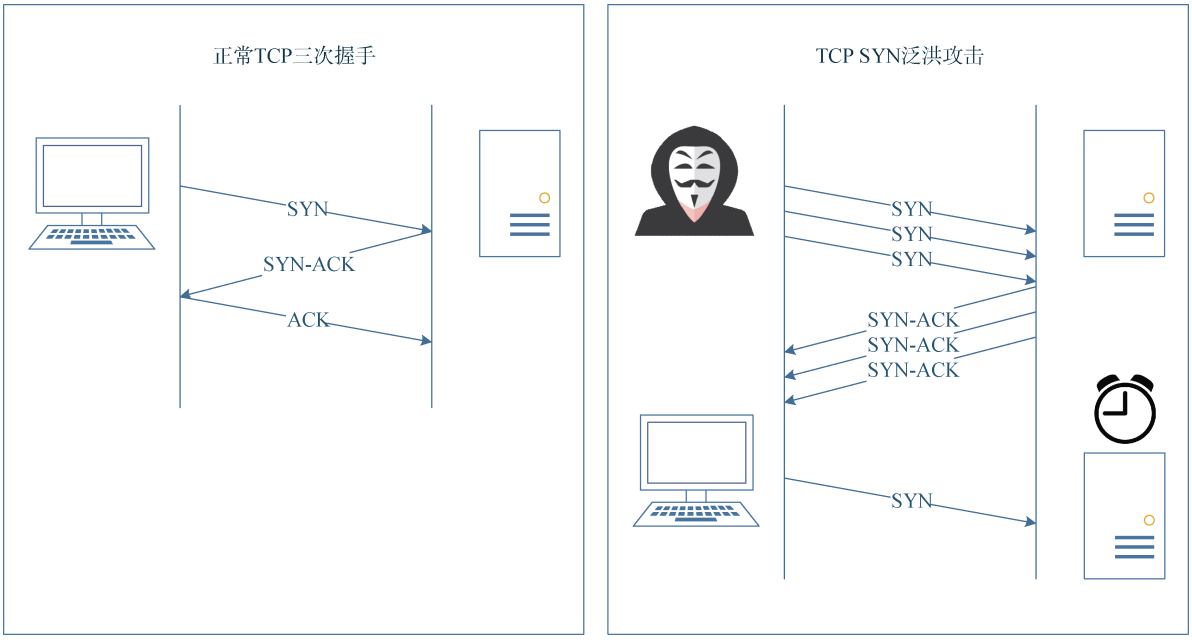

图 2　TCP 三次握手与 TCP SYN 泛洪攻击示意图

Fiuger 2　TCP three-way handshake and TCP SYN flood attack

过大量发送三次握手第一步 SYN 数据包请求, 服务器收到它们之后会回复 SYN-ACK, 随后等待用户发送 ACK 确认此次连接。然而恶意用户在此时并不会回复 ACK 连接导致目标节点一直会等待用户发送 ACK, 由此使目标节点的网络带宽等资源造成浪费。

UDP 泛洪攻击不同于 TCP SYN 攻击, 由于 TCP 协议是面向连接的, 在通信时双方需要提前建立好网络连接, 攻击者可以利用这个过程对目标用户进行攻击。而 UDP 协议面向非连接, 通信双方不必在进行通信前建立好连接, 所以攻击者无法从这个步骤进行攻击。由于计算机在收到数据包之后需要对数据包进行处理, 去包头, 处理数据等一部分操作, 而这些操作需要利用 CPU, 内存资源, 攻击者可以利用这一点对目标计算机进行攻击。如: 攻击者利用大量的僵尸网络同时或在一个较短的时间段内向目标用户发送大量的 UDP 短数据包。这样目标计算机在接收到这些数据包之后, 会试图利用 CPU 与内存资源处理它们, 但由于这些 UDP 数据包的数据巨大, 它们会耗尽目标计算机的 CPU, 内存与网络带宽等资源, 由此达成对目标计算机攻击的目的。这就是 UDP 泛洪攻击。

若上述攻击仅仅是一台计算机对另一台计算机攻击, 那么该攻击被称为 DoS 攻击, 即拒绝服务攻击。若上述攻击是攻击者控制多台计算机对目标节点进行攻击, 该攻击被称为 DDoS 攻击, 即分布式拒绝服务攻击。

3.3 信息熵

熵广义表示某系统的混乱程度。而在信息论领域里, 信息熵也称香农熵[26], 它最初由香农于 1948 年提出的, 用于研究传输信息中的信息量。定义离散随机变量 X 的信息熵为 X 变量信息量的数学期望, 如式(1)所示。其中, X 所有可能的取值为 $\{x_1,\cdots,x_n\}$。由数学期望和信息量的定义, (1)可以继续写成(2)所示形状。

$$H(X)=\mathrm{E}\left[\mathrm{I}(X)\right] \tag{1}$$

$$\begin{aligned} H(X)&=\sum_{i=1}^{n}\mathrm{P}(x_i)\mathrm{I}(x_i)\\ &=-\sum_{i=1}^{n}\mathrm{P}(x_i)\log_b \mathrm{P}(x_i)\end{aligned} \tag{2}$$

其中 $\mathrm{P}(x_i)$ 为 X 取值为 x_i 的概率。这里还需要说明的是, 当 log 的底数 b 取值不同时, $H(X)$ 的单位也不同。常用的有当 b 分别等于 2, e 和 10 时, 对应的单位分别为 bit, nat 和 ban。一般不加说明常取 2, 本文默认为 2, 单位为 bit。在实际使用时, X 取值 x_i 的先验概率 $\mathrm{P}(x_i)$ 利用频率近似。

3.4 随机森林

决策树是一种监督学习算法, 决策树学习的目的是为了产生一棵泛化能力强, 即处理未见示例能力强的决策树。但由于决策树常常出现过拟合的情况[27], 随机森林较好的解决了这一点。随机森林采用多个决策树进行投票, 通过该机制改变决策树的缺点。并且随机森林还有很多优点: 具有较好的准确率, 能够有效地运行在大数据集上, 能够评估各个特征在分类问题上的重要性, 在生成过程中, 能够获取到内部生成误差的一种无偏估计, 对于缺省值问题也能够获得很好得结果和几乎不用调参[28]。在训练随机森林中的决策树时, 需要对训练集进行随机有放回抽样训练。本文使用的决策树模型为 CART 决策树[29], 它在划分特征时利用基尼系数(Gini index)作为标准, 设数据集 D 的基尼系数为 $Gini(D)$, 则它的定义为(3)所示,

$$\begin{aligned} Gini(D)&=\sum_{k=1}^{K}p_k(1-p_k)\\ &=1-\sum_{k=1}^{K}p_k^2\\ &=1-\sum_{k=1}^{K}\left(\frac{|C_k|}{|D|}\right)^2\end{aligned} \tag{3}$$

其中, $|D|$ 为数据集 D 样本的个数, 数据集中有 K 个类别, 为 C_k, $k=1,2,3,\cdots,K$。$|C_k|$ 为第 k 个类别中样本的个数。显然(4)成立。特别的, 对于二分类问题, 有式(5)。

$$\sum_{k=1}^{K}|C_k|=|D| \tag{4}$$

$$Gini(D)=2p(1-p) \tag{5}$$

这里, p 任意一个类别所占的频度。对于基尼系数一个直观的解释是, 它直接反映了从数据集 D 中随机抽取两个样本, 其类别不一致的概率。因此, 当 $Gini(D)$ 越小时, 该数据集的纯度就越高。设特征 A 有 n 个不同取值 $\{a_1,a_2,\cdots,a_n\}$, 它们将数据集 D 划分成为 n 个子集 $D_1,D_2,\cdots,D_n$, 显然有(6), 其中 $|D_i|$ 为子集 D_i 中样本的数量。那么特征 A 的基尼系数定义为(7)。

$$\sum_{i=1}^{n}|D_i|=|D| \tag{6}$$

$$Gini(D\mid A)=\sum_{i=1}^{n}\frac{|D_i|}{|D|}Gini(D_i) \tag{7}$$

$$A_* = \arg\min_{A \in f} Gini(D \mid A) \tag{8}$$

它反映的就是特征 A 对数据集 D 中样本的区分能力, 值越小区分效果越优。于是我们在训练时挑选可以使(7)取值最小的特征作为最优划分特征, 即使(8)成立, 其中 f 为所有特征集合。

4 系统模型

模型会先采集正常网络数据与 DDoS 恶意网络数据并进行清洗融合。随后会读取融合数据集数据并利用随机森林算法训练检测分类器, 并利用该分类器对其他未知的网络数据进行分类判断。我们可以将系统部署在 eNB 上对所有流经 eNB 的网络数据进行监控。

由于 LTE-A 网络环境下的 DDoS 所涉及的设备众多, 本次实验以仿真形式进行。仿真是在 Ubuntu16.04 环境下搭建 3.28 版本 NS-3 仿真工具完成的, 选择使用的编程语言为 C++。实验仿真出 10 个 UE, 2 个 eNB, 每个 eNB 挂载 5 个 UE, 如图 3 所示。为更好逼近现实数据, 共仿真两次, 第一次随机挑选一些 UE 安装 DDoS 恶意应用, 这些应用会对核心网内实体如 S-GW, P-GW, MME 等实体进行 TCP SYN 和 UDP 泛洪攻击, 部署在 eNB 设备上的模型将对其进行数据采集。第二次在全部 UE 上安装正常应用, 模型也对其进行数据采集。随后给数据打上标签再进行训练。在最终检测数据时利用十折交叉验证法对模型性能进行评估。大体模型结构如图 4 所示。

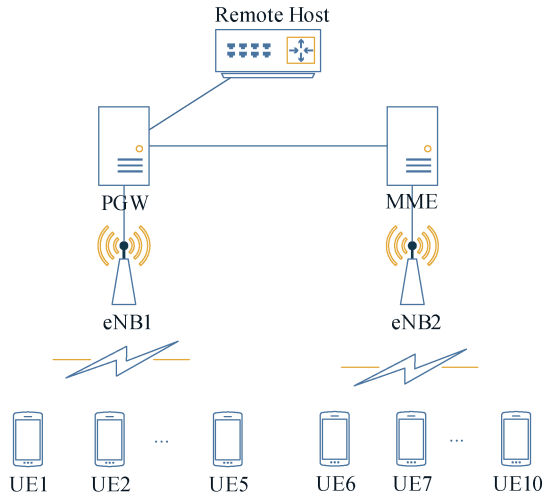

图 3 LTE 网络仿真架构

Figure 3 Architecture of LTE simulation

4.1 数据获取

在 NS-3 仿真系统中, 共仿真了两种 DDoS 攻击: 基于 TCP 协议的 SYN 协议攻击和基于 UDP 协议的泛洪攻击。正常传输数据有两种, 一种是基于 TCP 协议的数据包, 另一种是基于 UDP 协议的数据包。由于现实当中 LTE-A 网络中 UE 设备与各类应用数量较大, 网络当中的数据较为复杂, 为模拟出贴合实际网络中真实情况的数据传输, 正常传输数据包

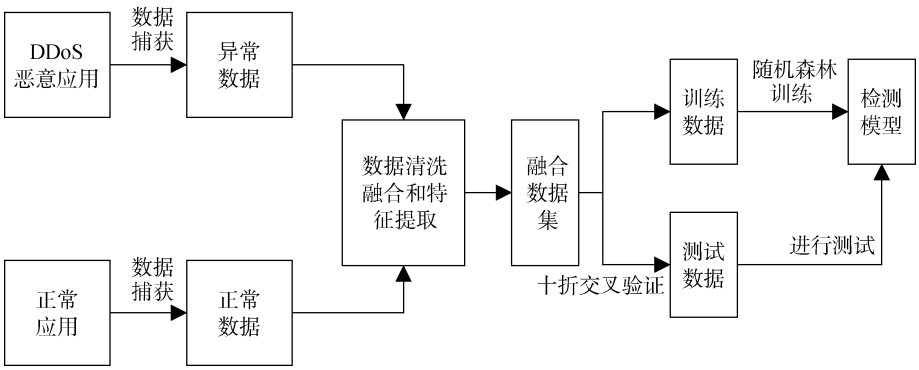

图 4 模型框图

Figure 4 Architecture of the model

的载荷长度是随机的。DDoS 恶意应用会大量发送 SYN 请求和短 UDP 数据包给核心网中的服务器或远程 PDN 当中的服务器试图耗尽 LTE-A 中的网络资源或利用 LTE-A 网络对远程服务器发起 DDoS 攻击。由于该模型使用的机器学习算法是随机森林算法, 为监督学习, 所以在获取数据时我们需要获取带标签数据。为实现这一点, 在仿真中, 共仿真两次, 第一次仿真的目的是得到 DDoS 攻击时的数据包并对数据打上攻击标签。具体实施方法是随机挑选部分 UE 安装 DDoS 恶意应用, 它们会随机 DDoS 泛洪攻击 LTE-A 网络中的核心网各实体或远程 PDN 当中的服务器, 这些应用会产生 DDoS 攻击数据, 这些攻击数据在途经 eNB 时会被部署在 eNB 上的模型捕获, 并打上攻击数据的标签。第二次仿真的目的是得到

正常通信时网络当中的正常数据并对数据打上正常的标签。具体实施方法是在各个 UE 上安装正常通信应用, 它们会正常与远端服务器进行通信, 主要利用的协议有 TCP 与 UDP 两种。这些应用会产生正常网络数据, 同理, 模型会捕获这些数据并打上正常数据的标签。在得到这些打好标签的数据之后, 模型就会对这些数据进行清洗, 融合以及特征提取, 以便之后训练模型使用。本次仿真共获取了 27784 个正常数据包, 4147 个攻击数据包。具体数据包数量如表 1 所示。

表 1 仿真数据数据数量

Table 1 Quantity of stimulation data

数据类型	协议类型	数量	共计
正常数据	TCP	22522	27784
	UDP	5262	
攻击数据	TCP	1427	4147
	UDP	2720	

4.2 数据清洗, 融合和特征提取

在 eNB 捕获网络数据并标记之后, 随后将对数据进行清洗, 特征提取和融合。仿真得到的数据为 pcap 文件, 利用 Python 编程语言的 Scapy 库对获取的网络数据文件进行处理。Scapy 库[30]是一个功能强大的交互式数据包操作工具。它能够伪造或解码大量协议的数据包, 通过线路发送, 捕获它们, 匹配请求和回复等。在数据特征提取方面, 需要对数据的各类特征进行提取或舍弃。一般来讲每个数据包的基本特征有: 时间, 序号, 所用协议, 原始 IP, 目的 IP, 原始端口, 目的端口和数据包载荷等。文献[15]中提出, 现实中很多数据包载荷内容在应用层被相关应用加密。所以本文在处理数据时只保留包头信息。需要说明的是, 在训练模型之前, 需要对非数值取值进行编码。例如特征 *protocol*, 取值为: TCP, UDP 等, 那么分别设 TCP=1, UDP=2 等。

我们除了在保留每个数据包的基本特征: 所用协议, 原始 IP, 目的 IP, 原始端口和目的端口等之外, 还要对数据集本身进行相应的处理以提取出新的特征。在本文提出的模型当中, 采用了我们提出的 2 个新特征, 即 *same_in_t_window* 和 *entropy_in_t_window*。这两个特征需要我们从数据集当中进行计算得出。我们分别对这两个特征进行介绍。

4.2.1 特征 *same_in_t_window* 介绍

特征 *same_in_t_window* 主要目的是统计时间窗口内的相同数据包数量。由于我们主要检测 DDoS 攻击, 而此类攻击的一大特征就是在某时间段之内, 网络中会出现大量的攻击数据, 而近几年主要的大流量 DDoS 攻击类型为 TCP SYN 泛洪攻击与 UDP 泛洪攻击, 这两种攻击类型的主要特征之一就是攻击者会在短时间内发送大量类似甚至相似数据。从这一特点出发, 我们利用 *same_in_t_window* 作为特征之一进行攻击识别。这里需要说明的是, 我们需要重新定义一下两数据包“相同”的条件: 若两数据包的源 IP, 目的 IP, 使用的协议均分别相同, 则称两数据包“相同”。考察一个数据包的时间窗口时, 具体时间区间为 [*t_current-t_window, t_current*], 若 *t_current-t_window*<0, 则区间为[0, *t_current*]。其中 *t_current* 为当前考察的数据包的时间戳, *t_window* 为当前的时间窗口大小, 该参数可以被调整, 单位为秒。根据上面的定义与说明, 我们可以对数据集进行相应的特征提取。

4.2.2 特征 *entropy_in_t_window* 介绍

特征 *entropy_in_t_window* 的主要目的是利用信息熵作为数据特征之一识别 DDoS 攻击。熵是衡量系统复杂程度的一个量, 信息熵可以衡量系统平均信息量的大小, 若系统当中平均信息量大, 则信息熵数值上会较大, 若系统当中平均信息量小, 则信息熵数值上会较小。在 LTE-A 网络当中未发生 DDoS 攻击正常通信时, 由于设备与应用众多, 某 eNB 在一个时间窗口内的全体数据包构成的数据集较为复杂, 它们包含的信息量会越大, 则该数据集合的信息熵会趋向于一个较高的水平。当网络中出现了 DDoS 泛洪攻击之后, 同前面的情况, 我们主要考虑 TCP SYN 泛洪攻击与 UDP 泛洪攻击这两种, 由于发生了此类攻击, 在攻击时, 某数据包在一定时间窗口内的所有网络数据构成的集合当中会存在大量相同的数据包, 而导致该集合内的信息量急剧减少, 则该集合的信息熵会趋向于一个较大值。由上面的分析, 我们发现信息熵这个量本身的特性可以帮助我们识别 DDoS 攻击。从这一特点出发, 我们利用 *entropy_in_t_window* 作为特征之一进行攻击识别。考察一个数据包的时间窗口时, 具体时间区间为 [*t_current-t_window, t_current*], 若 *t_current-t_window* <0, 则区间为[0, *t_current*]。数据包, 变量含义定义同 *same_in_t_window* 当中的定义。

4.2.3 特征选择

在对数据进行相应的清洗以及特征提取之后, 我们得到了 9 个特征, 这 9 个特征我们可以根据先前说明的那样计算各个特征的基尼系数。由于基尼系

数的值域为[0,1], 且越小表明特征对数据的分类效果越优, 我们对它进行放大处理, 均乘以 100。C4.5 决策树是取增益率作为特征提取时的指标, 与基尼系数类似, 增益率通过不同的计算公式得出, 也能反映出某个特征能否对数据集进行良好的划分, 与基尼系数不同的是, 增益率可以发现某些取值明显过多而导致可以很好的分类数据, 但这种分类方式是对我们结果完全无意义的分类。例如网络数据的时间戳或者是数据的序号, 由于每一条数据只对应一个时间戳或序号, 那么如果我们利用这两个特征来划分数据或训练模型, 得到的最终模型显然是无意义或是过拟合的。我们需要去掉这一类明显无意义的特征。所以我们不考虑数据包的时间, 序号和载荷这三个基本特征, 去掉它们之后再对模型进行训练。由于 LTE-A 网络当流经中某一个 eNB 的数据量巨大的, 所以训练模型时的时间消耗不容忽视。在保证准确率的前提下, 本次实验将采用基尼系数排名前三的特征进行模型训练。

由上述分析可以得到表 2, 其中, *s_ip*, *d_ip*, *s_port*, *d_port* 和 *length* 这些特征取值可以直接从模型捕获的数据包包头当中提取获得, 且数据本身不需要对其进行编码。虽然 *protocol* 与 *tcp_flag* 也可以从数据包当中获取, 但是由于这两个特征是离散“非数字”数据, 我们需要对他们进行相应的编码操作。如: 针对 *protocol* 这个特征, 可以设: TCP=1, UDP=2 等。针对 *tcp_flag* 这个特征, 在 TCP 协议当中包头中规定预留 8 位作为 TCP 的标志位, 每一位置 1 有效, 这样共有 2^8 种标志, 但是我们依次对它们进行从 1 开始的自然数编码会不够灵活并且可读性差, 我们采取表 3 所示的编码方式进行编码。如: TCP 标志为 SYN-ACK, 则 *tcp_flag* 取值为 2+16=18。其他情况同理。这样我们就会在不重复的前提下对所有所有 *tcp_flag* 的取值快速编码。若为非 TCP 协议如 UDP 协议, 则这个特征取值为–1。

表 2　提取的特征及其含义与基尼系数

Table 2　Features extracted and the meanings

特征	含义	Gini index ×100
protocol	数据包所用的协议。如: “*TCP*”等。	≈ 89
s_ip	数据包的发送端 IP。	≈ 130
d_ip	数据包的目的端 IP。	≈ 133
s_port	数据包的发送端端口号。	≈ 69
d_port	数据包的目的端端口号。	≈ 133
tcp_flag	TCP 的标志位。如果是非 TCP 协议, 取值–1, 否则按照表 3 取值。	≈ 89
length	数据包长度。	≈ 0.34
same_in_t_window	在当前数据包的周围时间窗口内拥有于当前数据包“相同”的数据包数量。具体的时间区间为: [*t_current-t_window, t_current*], 其中, *t_current* 为当前数据包的时间, *t_window* 为时间窗口大小。需要说明的是, 判断两数据包是否相同方法如下: 若两数据包的源 IP, 目的 IP, 数据所使用的协议分别相同, 则称两数据包“相同”。	≈ 47
entropy_in_t_window	在当前数据包的周围时间窗口内的信息熵。具体的时间区间为: [*t_current-t_window, t_current*], 参数定义同上, 在计算信息熵时, 也需要定义两个数据包是否“相同”, 定义同上。	≈ 0.48

表 3　*tcp_flag* 取值表

Table 3　Values of *tcp_flag*

TCP 标志位	*tcp_flag*
(非 TCP 协议)	–1
FIN	1
SYN	2
RST	4
PSH	8
ACK	16
URG	32
ECE	64
CWR	128

注: 若 TCP 标志为 SYN-ACK, 则 *tcp_flag* 取值为 2+16=18。其他情况同理。

在得到所有的特征之后, 我们就可以利用前面对一个特征的基尼系数的定义来计算所有特征的基尼系数。分别对特征进行计算之后, 由于基尼系数的值域在[0,1], 且部分重要特征的值非常接近 0, 我们对其进行放大, 增加其可读性。将计算出来的值全部乘 100, 不改变它们的排名情况, 注意, 基尼系数越小说明该特征对数据集划分效果越好。根据计算可以看出, 基尼系数排名三个特征分别为: *length*, *same_in_t_window* 和 *entropy_in_t_window*。

4.3　融合数据集

在数据清洗, 融合和特征提取之后, 由于仿真时是分两次得到的正常数据与攻击数据, 我们需要对两个数据集合进行数据融合, 得到一个总体数据

集。具体融合的方法为: 根据两个数据集当中数据的时间戳按照顺序依次排列, 得到最终的融合数据集, 如图 5 所示, 其具体过程类似于融合两个有序表, 得到一个新的有序表。

4.4 抽取测试数据

在经过了随机森林训练之后, 我们得到了模型分类器, 此时需要对分类器的识别准确率等指标做出评估测试, 因此需要在融合好的数据集当中, 抽取出一定量的有标签数据对训练好的模型准确率等进行测试。本次实验采用十折交叉验证(10-fold cross validation)法, 即将数据集均为 10 份, 轮流将其中 9 份作为训练数据, 1 份作为测试数据, 第一次使用前 9 份作为训练数据, 第 10 份作为测试数据, 第二次使用第 2 份作为测试数据, 其余作为训练数据, 以此类推。10 次结果的各指标的均值作为算法各指标的评估结果。

4.5 训练模型分类器

在得到训练数据集与测试数据集之后, 就可以用随机森林算法对模型分类器进行训练。本次实验利用 scikit-learn[37]作为工具训练随机森林模型分类器。作为一个简单高效的数据挖掘和数据分析工具, scikit-learn 可以实现数据预处理, 分类, 回归, 降维等操作, 并且开源, 可以供大家共同研究, 开发和维护。在 scikit-learn 工具中, 可以使用 RandomForestClassifier()函数训练随机森林分类模型, 其中参数设置如表 4 所示, 其中很多参数采取默认取值。可以看到, 训练出的随机森林当中存在 10 棵最大深度为 3 的决策树, 这既保证了较高的识别准确率, 也保证了较低的识别耗时。但由于训练模型时需要用到本文提出的两种特征, 而这两种特征在实际操作时均需要较长的耗时, 所以训练过程本身的耗时会较高, 但是一旦训练好模型之后, 其识别时耗时非常短, 具体情况将会在后面详细说明。

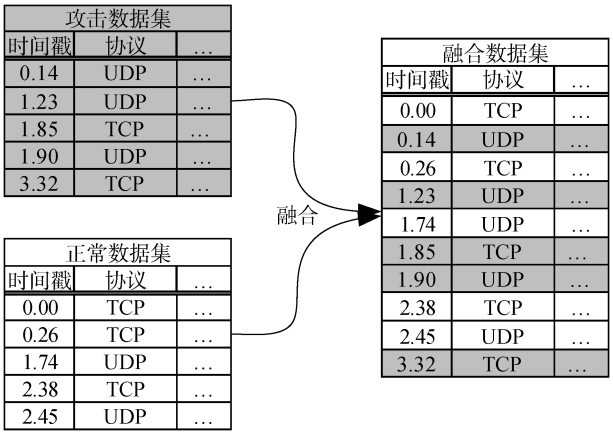

图 5 正常数据与攻击数据融合示意图

Figure 5 The fusion of normal data and malicious data

表 4 RandomForestClassifier()函数参数

Table 4 Arguments of RandomForestClassifier()

参数名称	含义	取值
bootstrap	训练随机森林时是否采用有放回式的抽样。	TRUE
class_weight	分类权重, 若为空, 则所有类别权重均为 1。	None
criterion	采用何种方法评估参数分类质量。	'gini'
max_depth	随机森林当中每棵决策树的最大深度。	3
max_features	寻求最佳分割时的考虑的特征数量, 即特征数达到多大时进行分割。	'auto'
max_leaf_nodes	最大叶子节点数, 以最好的方式生成树。最好的节点定义为与大多数分类不相同数据数量相对较少, 即纯度较高的叶子节点。 如果为 None 则无限叶子节点数量。	None
min_samples_leaf	叶子节点上包含的样本最小值。	1
min_samples_split	分割节点所需的最少样本数量, 即超过该数量进行分裂。	2
min_weight_fraction_leaf	定义能够成为叶子节点的条件。能成为叶子节点的条件是: 该节点对应的实例数和总样本数的比值, 大于等于 min_weight_fraction_leaf 的值。	0
n_estimators	森林中决策树的个数。	10
n_jobs	拟合和预测过程中并行运用的作业数量。如果为–1, 则作业数设置为处理器的内核数。	1
oob_score	使用 oob 数据集测试得到的得分数。	FALSE
random_state	该参数定义了随机数产生器使用的随机数种子。如果为 None, 则随机数生成器使用 np.random 的 RandomState 实例进行随机数生成。	FALSE
verbose	控制拟合和预测时的冗余度。	0
warm_start	若为 TRUE, 则重用上一个调用的解决方案以适合并向整体添加更多估算器, 否则只需适合整个新林。	FALSE

4.6 测试模型

如 4.4 节所述, 本次采用十折交叉验证法对模型的准确率, *TP*, *FP*, *FN* 和 *TN* 进行验证。其中, *TP* 代表真正例, 分类正确, 把原本属于正类的样本分成正类。*FP* 代表假正例, 分类错误, 把原本属于反类的错分成了正类。*FN* 代表假反例, 分类错误, 把原本属于正类的错分成了反类。*TN* 代表真反例, 分类正确, 把原本属于反类的样本分成反类, 如表 5 所示。可以看出, 准确率可以通过(9)计算。

表 5 指标说明

Table 5 Meaning of measures

真实情况	预测结果	
	正例	反例
正例	*TP*(真正例)	*FN*(假反例)
反例	*FP*(假正例)	*TN*(真反例)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

除了上述参数之外, 我们还测试了两个时间窗口对准确率与模型训练耗时的影响, 分别取 *same_in_t_window* 与 *entropy_in_t_window* 在 1,2,3,⋯,10 当中的所有排列组合取值组合, 并依次训练模型, 并对模型的准确率以及训练耗时进行评估与统计, 在保证准确率的情况下尽可能的缩短系统训练时间提高系统性能。

5 测试结果与分析

由于 NS-3 仿真器内部机制的原因, 我们无法查看被攻击节点的 CPU 与内存的占用情况, 丢包率也无法良好的进行测试, 即无法查看被 DDoS 泛洪攻击目标的系统资源, 包括 CPU, 内存和网络带宽资源的消耗情况, 所以仅仅查看模型对于 DDoS 攻击行为的数据包的检测情况。如上面所述, 十折交叉验证为将数据集均为 10 份, 轮流将其中 9 份作为训练数据, 1 份作为测试数据, 10 次各指标结果的均值作为算法各指标的评估。

5.1 模型训练耗时分析

首先考察本文提出的两个时间窗口参数 *same_in_t_window* 与 *entropy_in_t_window* 对系统训练耗时的影响。由于训练耗时很长, 先进行两参数的粗颗粒度搜索, 两参数取遍 1 至 10 整数间的所有排列组合, 在此只进行单独一次训练得到准确率结果而不选择十折验证即训练十次。其结果如图 6 所示。其中 e_size 与 t_size 两参数分别代表: *entropy_in_t_window* 与 *same_in_t_window*。从图中可以看到, 模型训练耗时几乎只随 *entropy_in_t_window* 的增加而增加。而当 *entropy_in_t_window* 取值相近的情况下, 模型检测准确率曲线走势大体随 *same_in_t_window* 起伏。且这两参数在取值均较高时可以明显的提升模型准确率, 但是带来的弊端也很明显, 即训练耗时明显增加, 并有可能出现过拟合的情况。过拟合即模型只对本训练数据集合的准确率等性能良好, 但是遇到新的数据或数据集的性能欠佳, 这是由于在我们在训练模型时过度苛求准确率而不择手段的提高它, 可能导致模型只会分辨出当前数据集当中出现过的攻击数据, 而对新的攻击数据无法分辨。为了避免这种情况, 可以适当的允许一些错误的出现而不能只追求准确率。我们从图中还可以看到, 两参数在取值较低的情况下对模型准确率的影响较小。在训练耗时与准确率二者之间, 需要进行一些折中。

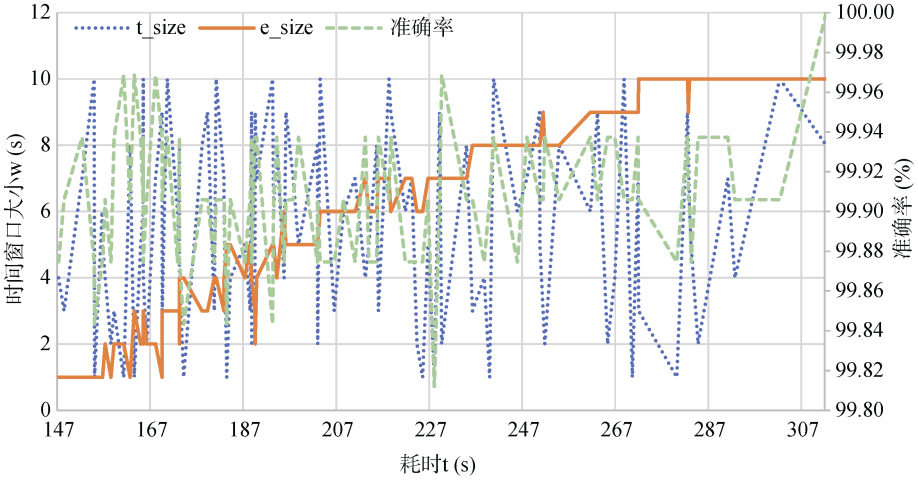

图 6 各时间窗口取值与耗时图表

Figure 6 Size and corresponding time consumption of each time window

上述现象表明: 本文提出的两个参数对模型准确率均有影响, 且其中信息熵的时间窗口即 *entropy_in_t_window* 严重影响模型训练耗时。当二者取值均较高时, 模型可以得到较好的准确率, 但是耗时可能会严重增加并且可能出现过拟合的情况。

5.2 模型分类耗时分析

需要说明的是, 算法训练时需要在数据集当中动态更新时间戳的取值范围即需要统计的时间戳范围, 并在附近数据当中查询并统计相同的数据数量, 维护一个需要统计数据包的数组, 要消耗大量的时间, 这使得训练一个模型也需要消耗大量的时间。但模型当中的随机森林由 10 棵决策树构成, 并且单棵决策树的深度只有 3, 那么设一个新数据在决策树当中与一个节点进行比较并走向孩子节点的耗时为 τ_{comp}, 那么模型在得到一个新数据并得到该数据的预测标签只需要比较 20 次, 耗时为 $20\tau_{comp}$。即当模型得到一个新数据并得到该数据的分类标签需要的时间仅为 $20\tau_{comp}$, 而注意到 τ_{comp} 只是计算机进行一次条件比较的耗时, 现实当中这个值会非常小以至于可以忽略不计, 则我们可以看到, 虽然训练出来该模型需要消耗大量的时间, 但是在训练好模型之后部署在 eNB 上投入使用之后, 在数据包流中甄别出 DDoS 泛洪攻击的数据包所需的时间很短。

5.3 适当选取参数进行十折验证

由于计算信息熵时需要消耗大量的时间, 我们选取适当的参数取值进行十折验证。在不损失过多的准确率的情况下尽可能减少训练耗时, 选择 *entropy_in_t_window*=2, *same_in_t_window*=1, 进行十折交叉验证, 验证结果如表 6 所示, 表中的所有结果均为 10 次训练测试出的结果均值。可以看出, 选取该组参数取值准确率较高, 约为 99.956%, *TP*, *TN* 均为判断准确的数据数目, 而 *FP* 为假正例, 即遗漏未被模型检测出的攻击数据, 10 次平均为 1.1 个。而 *FN* 为假反例, 即原本正常却被模型误报为异常的数据, 10 次平均为 0.3 个。模型训练耗时平均为 161.33 秒。如此选取参数取值, 在保证了基本的准确率之外还尽可能的减少了系统模型训练所需的时间。

表 6 十折交叉验证结果

Table 6 Results of 10-fold cross validation

指标	十次平均结果
准确率	99.956%
TP	2918.8 个
FP	1.1 个
FN	0.3 个
TN	270.9 个
耗时	161.33s

综上所述, 利用信息熵作为特征之一训练的基于随机森林的 LTE-A 网络 TCP SYN 与 UDP 的 DDoS 检测模型对于检测网络当中 TCP SYN 与 UDP 恶意 DDoS 流量具有较高的准确率以及较低的误报率与漏检率。但是模型训练耗时较高, 与误报数据相比, 模型更容易漏检攻击数据, 而误报的情况相对较少。且它不仅可以检测出 UE 僵尸网对核心网重要部件的泛洪攻击, 也可以检测出流经 LTE-A 网络当中 UE 僵尸网对于远程服务器的泛洪攻击。

6 小结

LTE-A 作为全 IP 异构架构网络, 极易遭受 DDoS 网络攻击。据相关学者调查, 2017 年在大流量网络攻击当中 TCP SYN 泛洪攻击与 UDP 泛洪攻击占相当大的比重, 虽然 5G 相关标准与技术已被确定或尝试投入使用, 但短时间内我们依然无法离开 LTE-A 网络架构并且它依然会在一段时间内为我们提供服务。在其他无线网比如 IoT, Ad-hoc 甚至 IEEE 802.15.4 均出现了 DDoS 攻击, LTE-A 也易遭受多种 DDoS 攻击方法。而针对利用 LTE-A 当中 UE 僵尸网对 LTE-A 核心网或远程服务器进行 DDoS 攻击的研究较少。鉴于此, 本文提出了一种针对 LTE-A 网络中的 DDoS 攻击流量检测模型, 模型利用本文提出的两个新特征 *entropy_in_t_window*, *same_in_t_window* 可以有效的检测出 LTE-A 网络中的 TCP SYN 与 UDP 泛洪 DDoS 攻击流量。该模型的优点有: 检测耗时短, 准确率较高, 达 99.956%, 误报率与漏检率均较低, 部署在 eNB 上的检测模型不仅可以检测出 UE 僵尸网对 LTE-A 核心网中的重要部件的 DDoS 攻击, 还能有效的检测出利用 UE 僵尸网对远程服务器的 DDoS 攻击。模型的缺点有: 仅仅对此两种 DDoS 攻击有良好的识别能力。模型训练需要的时间也很长, 不够灵活。

7 未来工作

由于 DDoS 攻击的种类较多, 带来的攻击也很大, 本文提出的模型只对两种攻击具有良好的识别能力, 所以今后重点将会放在研究能识别更多种类的 DDoS 攻击上。并且本文提出的模型在训练时会消耗很多的时间, 在真正投入使用时会带来一些不便, 以后的工作也要继续研究优化模型, 减少训练时间。并在 NS-3 中仿真更多种类的 LTE-A 攻击, 试图利用不同的机器学习算法对这些攻击进行识别与防御。未来还将会把更多精力投入到 LTE-A 信令攻击并具体硬件实现当中去。

未来 5G 网络会更加的普及, 取代现有很多网络成为我们身边不可或缺的一部分也需要将目光投入到 5G 网络安全中, 试图利用机器学习相关算法检测网络当中的信令攻击, 或直接设计协议分析模型对 5G 网络当中的协议可能出现的漏洞进行检测分析。

致　谢　本论文由国家重点研发计划项目(No.2016YFB0800700)与国家自然科学基金项目(No.61772404, No.61602359)资助。

参考文献

[1] "LTE-Advanced," Jeanette Wannstrom, http://www.3gpp.org/technologies/keywords-acronyms/97-lte-advanced, Jun. 2013.

[2] "The State of LTE," OpenSignal, https://opensignal.com/reports/2017/06/state-of-lte, Jun. 2017.

[3] A. Sardana, and R. Joshi, "An auto-responsive honeypot architecture for dynamic resource allocation and QoS adaptation in DDoS attacked networks," *Computer Communications*, vol. 32, no. 12, pp. 1384-1399, Jul. 2009.

[4] "DISTRIBUTED DENIAL OF SERVICE (DDOS) WHAT DOES DDOS MEAN?" Incapsula, https://www.incapsula.com/ddos/ denial-of-service.html.

[5] 绿盟科技发布《2017 上半年 DDoS 与 Web 应用攻击态势报告》, 绿盟科技, http://blog.nsfocus.net/2017-mid-year-ddos-web- cybersecurity-threat-report/, Aug. 2017.

[6] C Kolias, G Kambourakis, A Stavrou, and J Voas, "DDoS in the IoT: Mirai and other botnets," *Computer*, vol. 50, no. 12, pp. 80-84, Jul. 2017.

[7] J. Henrydoss, and T. Boult, "Critical security review and study of DDoS attacks on LTE mobile network," IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), pp. 194-200, Oct. 2014.

[8] L. Qiang, W. Zhou, B. Cui, and L. Na, "Security analysis of TAU procedure in LTE network," Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), pp. 372-376, Nov. 2014.

[9] R. Bassil, I. H. Elhajj, A. Chehab, and A. Kayssi, "Effects of signaling attacks on LTE networks," Workshops of 27th International Conference on Advanced Information Networking and Applications (WAINA), pp. 499-504, Mar. 2013.

[10] R. Bassil, A. Chehab, I. Elhajj, and A. Kayssi, "Signaling oriented denial of service on LTE networks," ACM international symposium on Mobility management and wireless access (MobiWac 2012), vol. 8, no. 4, pp. 153-158, Oct. 2012.

[11] M Khosroshahy, D Qiu, and MKM Ali, "Botnets in 4G cellular networks: Platforms to launch DDoS attacks against the air interface," International Conference on Selected Topics in Mobile and Wireless NETWORKING (MoWNet'13), vol. 143, no. 6, pp. 30-35, Sept. 2013.

[12] L He, Z Yan, M Atiquzzaman. "LTE/LTE-A Network Security Data Collection and Analysis for Security Measurement: A Survey," *IEEE Access*, vol. 6, pp. 4220-4242, Jan. 2018.

[13] R Doshi, N Apthorpe, and N Feamster. "Machine Learning DDoS Detection for Consumer Internet of Things Devices," *arXiv preprint arXiv:1804.04159*, Apr. 2018.

[14] C Livadas, R Walsh, D Lapsley, and WT Strayer, "Using machine learning technliques to identify botnet traffic," IEEE Conference on Local Computer Networks (LCN), pp. 967-974, Nov. 2006.

[15] A Feizollah, NB Anuar, R Salleh, F Amalina, RR Ma'arof, and S Shamshirband, "A study of machine learning classifiers for anomaly-based mobile botnet detection," *Malaysian Journal of Computer Science*, vol. 26, no. 4, pp. 251-265, Dec. 2013.

[16] JH Jun, H Oh, and SH Kim, "DDoS flooding attack detection through a step-by-step investigation," Networked Embedded Systems for Enterprise Applications (NESEA), pp. 1-5, Dec. 2011.

[17] L Feinstein, D Schnackenberg, R Balupari, and D Kindred, "Statistical approaches to DDoS attack detection and response," *IEEE Xplore*, vol. 1, pp. 303-314, Apr. 2003.

[18] S Yu, W Zhou, R Doss, and W Jia "Traceback of DDoS attacks using entropy variations," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 22, no. 3, pp. 412-425, Mar 2011.

[19] K Kumar, RC Joshi, and K Singh. "A distributed approach using entropy to detect DDoS attacks in ISP domain," International Conference on Signal Processing (ICSPIS), pp. 331-337, Nov. 2007.

[20] A Lakhina, M Crovella, and C Diot, "Mining anomalies using traffic feature distributions," *In ACM SIGCOMM Computer Communication Review (ACM)*, vol. 35, no. 4, pp. 217–228, Oct. 2005.

[21] X Ma, and Y Chen, "DDos detection method based on chaos analysis of network traffic entropy," *IEEE Communications Letters*, vol. 18, no. 1, pp. 114–117, Dec. 2014.

[22] J´erôme Franc¸ois, Issam Aib, and Raouf Boutaba. "Firecol: a collaborative protection network for the detection of flooding ddos attacks," *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 6, pp. 1828–1841, Dec. 2012.

[23] Y Tao and S Yu, "DDos attack detection at local area networks using information theoretical metrics," In Trust, *Security and Privacy in Computing and Communications (TrustCom),* pp. 233–240, 2013.

[24] A. Kesavan, "Three Types of DDoS Attacks," ThousandEyes Blog, https://blog.thousandeyes.com/three-types-ddos-attacks/, Nov. 2016.

[25] R Balian, "Entropy, a Protean concept," Poincaré Seminar, *Progress in Mathematical Physics*, pp. 119–144, Nov. 2004.

[26] T Hastie, R Tibshirani, JH Friedman, and J Franklin, "The Elements of Statistical Learning (2nd ed.)," *Springer*. ISBN, vol.27, no. 2, pp.587-588, 2008.

[27] "[Machine Learning & Algorithm]随机森林(Random Forest)," Poll, http://www.cnblogs.com/maybe2030/p/4585705.html, Jun. 015.

[28] Breiman L. "Classification and regression trees," Routledge, 2017.

[29] "Scapy", Philippe Biondi and the Scapy community, https://scapy.net/, 2018.

[30] C Kolias, G Kambourakis, A Stavrou, and J Voas, "DDoS in the IoT: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp.80-84, Jul. 2017.

[31] SA Arunmozhi, and Y Venkataramani. "DDoS Attack and Defense Scheme in Wireless Ad hoc Networks," *arXiv preprint*

arXiv:1106.1287, vol. 3, no. 3, Jun 2011.

[32] P Sharma, N Sharma, R Singh, "A Secure Intrusion detection system against DDOS attack in Wireless Mobile Ad-hoc Network," *International Journal of Computer Applications*, vol. 41, no. 21, pp. 7-14, Mar. 2012.

[33] C Balarengadurai, and S Saraswathi, "Comparative analysis of detection of DDoS attacks in IEEE 802.15. 4 low rate wireless personal area network". *Procedia engineering*, vol. 38, no. 1, pp. 3855-3863, Jun. 2012.

[34] D Rupprecht, K Kohls, T Holz, and C Pöpper, "Breaking LTE on Layer Two," *IEEE Computer Society*, May. 2019.

[35] A Gupta, T Verma, S Bali, and S Kaul, "Detecting MS initiated signaling DDoS attacks in 3G/4G wireless networks," *Communication Systems and Networks*, pp.1-60, Feb. 2013.

[36] M Khosroshahy, D Qiu, and MKM Ali, "Botnets in 4G cellular networks: Platforms to launch DDoS attacks against the air interface," In Selected Topics in Mobile and Wireless Networking (MoWNeT), vol. 143, no. 6, pp.30-35. Sept. 2013.

[37] "3.2.4.3.1. sklearn.ensemble.RandomForestClassifier," scikit-learn community, http://scikit-learn.org/dev/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier.

龚宇翔 陕西西安人，西安电子科技大学硕士研究生，主要研究方向为大数据，机器学习，LTE/LTE-A/5G 安全与隐私保护。Email: gyx215@outlook.com

曹进 陕西西安人，博士，西安电子科技大学副教授，硕士生导师。主要研究方向为应用密码学，安全协议分析，无线网络安全。Email: jcao@xidian.edu.cn

付玉龙 黑龙江齐齐哈尔人，博士，西安电子科技大学讲师，硕士生导师。主要研究方向为逻辑代数与形式化安全证明，5G 网络安全，安全数据采集。Email: ylfu@xidian.edu.cn

郭敏 山东潍坊人，硕士，工程师，主要研究方向为网络安全。Email: guominjmh@163.com