SDN 下基于深度学习混合模型的 DDoS 攻击检测与防御

李传煌，吴艳，钱正哲，孙正君，王伟明
（浙江工商大学信息与电子工程学院，浙江 杭州 310018）

摘 要：软件定义网络（SDN, software defined network）作为一种新兴的网络架构，其安全问题一直是 SDN 领域研究的热点，如 SDN 控制通道安全性、伪造服务部署及外部分布式拒绝服务（DDoS, distributed denial of service）攻击等。针对 SDN 安全中的外部 DDoS 攻击问题进行研究，提出了一种基于深度学习混合模型的 DDoS 攻击检测方法——DCNN-DSAE。该方法在构建深度学习模型时，输入特征除了从数据平面提取的 21 个不同类型的字段外，同时设计了能够区分流类型的 5 个额外流表特征。实验结果表明，该方法具有较高的精确度，优于传统的支持向量机和深度神经网络等机器学习方法，同时，该方法还可以缩短分类检测的处理时间。将该检测模型部署于控制器中，利用检测结果产生新的安全策略，下发到 OpenFlow 交换机中，以实现对特定 DDoS 攻击的防御。

关键词：分布式拒绝服务；软件定义网络；攻击检测；深度学习
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2018128

DDoS attack detection and defense based on hybrid deep learning model in SDN

LI Chuanhuang, WU Yan, QIAN Zhengzhe, SUN Zhengjun, WANG Weiming
School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

Abstract: Software defined network (SDN) is a new kind of network technology, and the security problems are the hot topics in SDN field, such as SDN control channel security, forged service deployment and external distributed denial of service (DDoS) attacks. Aiming at DDoS attack problem of security in SDN, a DDoS attack detection method called DCNN-DSAE based on deep learning hybrid model in SDN was proposed. In this method, when a deep learning model was constructed, the input feature included 21 different types of fields extracted from the data plane and 5 extra self-designed features of distinguishing flow types. The experimental results show that the method has high accuracy, it's better than the traditional support vector machine (SVM) and deep neural network (DNN) and other machine learning methods. At the same time, the proposed method can also shorten the processing time of classification detection. The detection model is deployed in SDN controller, and the new security policy is sent to the OpenFlow switch to achieve the defense against specific DDoS attack.

Key words: distributed denial of service, software defined network, attack detection, deep learning

收稿日期：2018−02−28；**修回日期**：2018−05−16
基金项目：国家重点研发计划基金资助项目（No.2017YFB0803202）；浙江省自然科学基金资助项目（No.LY18F010006）；浙江省新型网络标准与应用技术重点实验室基金资助项目（No.2013E10012）；浙江省重点研发计划基金资助项目（No.2017C03058）
Foundation Items: The National Key Research and Development Program of China (No.2017YFB0803202), The Natural Science Foundation of Zhejiang Province (No.LY18F010006), The Key Laboratory of New Network Standards and Technologies of Zhejiang Province (No.2013E10012), The National Key Research and Development Program of Zhejiang Province (No.2017C03058)

1　引言

分布式拒绝服务（DDoS, distributed denial of service）[1]攻击是一种具有极强危害性的分布式、大范围协同作战的网络攻击方式，攻击者利用其控制的众多傀儡机，同时向被攻击目标发起拒绝服务（DoS, denial of service）攻击，最终导致被攻击目标的系统资源耗尽甚至崩溃，被攻击目标“拒绝”为正常用户提供所需服务。DDoS 攻击主要针对被攻击目标的系统资源和网络带宽，攻击范围包括网络层到应用层。自 1999 年发生第一起 DDoS 攻击以来，DDoS 已经成为广泛且致命的网络安全威胁之一。根据 Radware 公司的调查报告显示，DDoS 攻击是目前 Internet 相关组织所面临的最大的网络安全威胁[2]。内容传送网络（CDN, content delivery network）服务提供商 Akamai 的互联网安全状况报告显示，2017 年第四季度与 2016 年第四季度相比，DDoS 攻击总数增加了 14%，这也表明 DDoS 攻击长期处于总体上升趋势[3]，DDoS 攻击已经对系统和网络造成了严重的安全威胁。

软件定义网络[4]作为一种新型的网络架构，其核心思想是将网络设备的数据转发与决策控制功能进行分离，实现硬件的集中式控制。随着 SDN 应用的普及，SDN 的安全性问题已经成为 SDN 领域关键的研究课题之一。在基于 OpenFlow 技术的 SDN 中，较常用的 DDoS 攻击防御架构使用 OpenFlow 交换机采集网络流量，并解析数据分组的特征值信息，然后与 DDoS 攻击规则库进行规则匹配[5]，最终利用控制器完成入侵响应。其中，OpenFlow 交换机除了需要完成流量转发任务外，还需要完成数据协议分析和 DDoS 攻击规则库匹配等额外任务，控制器则除了需要完成维护及控制转发等常规任务外，还需要进行 DDoS 攻击数据分组特征收集和入侵响应[6]等特殊任务。控制器和交换机所带来的大量额外任务均会使那些已经负担过重的网络设备雪上加霜。因此，针对 SDN 体系架构的特点，构建一种高效合理的 DDoS 防御机制，是在设计和部署整个 SDN 架构时必须慎重考虑的关键性问题。

DDoS 检测是 DDoS 主要的防御机制之一，因为在大多数情况下，攻击流量与合法流量非常相似，攻击者会尝试模仿 Flash 群，导致 DDoS 攻击较难被自动检测，而流量不足的攻击行为甚至可以被视为一个早期的合法行为。许多研究人员尝试使用统计机器学习方法来检测 DDoS 攻击并丢弃攻击数据分组。传统机器学习方法对基于统计特征的 DDoS 攻击进行分类，性能优于统计方法。但该方法仍然有如下缺点：需要广泛的网络专业知识和 DDoS 的实验来选择合适的统计特征；仅限于一个或几个 DDoS 攻击向量；需要更新其模型和阈值来满足系统和攻击向量的变化；易受到低攻击速率影响。

深度学习为解决传统机器学习的局限性提供了可能。深度学习算法是一个对特征学习的过程，在学习过程中能够发现多层特征，并将高层特征表示成更抽象的数据特征。目前，深度学习已经被 Google、Facebook、Microsoft[7]、百度[8]等公司在语音识别、计算机视觉、自然语言处理等领域广泛应用。

本文提出了一种 SDN 中基于深度学习混合模型的 DDoS 攻击检测方法——DCNN-DSAE，该方法将 SDN 与深度学习相结合，并利用深度卷积神经网络（DCNN, deep convolution neural network）和深度堆栈自编码（DSAE, deep stacked autoencoder）进行检测。DCNN-DSAE 混合模型汲取了 DCNN 和 DSAE 算法的优点，不仅可以提高分类检测的精度，还可以缩短分类检测的处理时间，有效地缓解了 SDN 中的 DDoS 攻击效应，从而避免了网络资源的耗尽。

2　研究现状

DDoS 攻击如今已经成为网络安全方面的严重威胁之一[9-10]。随着互联网技术及应用的发展，DDoS 攻击出现的次数正在大幅增长，一个主要原因就是僵尸网络的出现和发展。这些僵尸网络是由恶意软件或机器组成的攻击网络，攻击者利用恶意流量不断地攻击受害者的服务器，使服务器不能为用户提供正常的服务，甚至导致网络瘫痪。

到目前为止，国内外针对 DDoS 攻击安全防御问题的主要解决方法是实时的网络监控，当 DDoS 攻击发生时，启动攻击流量清洗设备屏蔽 DDoS 攻击源，从而避免网络遭受 DDoS 攻击的侵害，达到安全防御的目的[11]。这个过程主要包括以下几点：在交换机、路由器等网络转发设备中进行源 IP 地址的合法性确认，建立黑/白名单；基于统计学方法，根据网络流密度变化情况，对网络流量进行实时监控；在路由转发设备中，建立源地址和转发设备对

应的入端口的映射表，通过对比数据流源 IP 地址与相应入端口地址，确定当前网络是否有 DDoS 攻击行为，然后对 DDoS 攻击流量进行清洗。

Mousavi 等[12]提出了一种在 SDN 控制器中通过计算熵值来检测 DDoS 攻击的入侵检测系统。该入侵检测系统的检测精度取决于熵的阈值，然而阈值的选择是通过调整参数大小的实验得到的，该方法具有一定的不可靠性。

Wang 等[13]提出了一种 SDN 中基于熵的 DDoS 攻击检测系统，通过处理交换机中流表的统计信息，达到检测系统中防止 DDoS 攻击的目的。虽然该方法减少了控制器在收集流表统计信息时的开销，但该方法试图在交换机等设备中增强智能控制功能，与 SDN 转发和控制分离的核心思想矛盾。

Jadidi 等[14]提出了一种基于流的异常检测系统，主要采用基于多层感知器和重力的搜索算法。该系统可以很好地区分正常类型的数据流和单一攻击类型的数据流，但其对组合攻击类型的数据流区分度较差。

Winter 等[15]提出了一种基于支持向量机（SVM, support vector mechine）算法的网络入侵检测系统，该方法具有较低的误报率。然而该系统在训练模型时只采用了攻击类型的数据训练模型，没有涵盖正常网络流量数据集，存在一定的片面性。

Trung 等[16]结合硬检测阈值和模糊推理系统（FIS），根据正常和攻击状态下的实际流量特征来检测 DDoS 攻击，但他们只考虑了分布时间、每个流的数据分组数量及分配到服务器的流量等几个特征，特征信息量不足。

此外，本文作者[17-18]曾将深度学习与 DDoS 检测相结合，使用深度学习方法直接对通过的流量本身进行检测，具有较高的精确度。由于该方法未对流量的统计特征进行检测，与轻量级的统计特征检测不同，它不能直接部署在 SDN 控制器中。

综上所述，现有的 DDoS 攻击检测方法较多，但它们仍存在着检测时延长、检测精度低、误报率较高、对新型 DDoS 攻击检测能力较弱等诸多问题。本文针对上述问题，并结合深度学习，提出了一种基于深度学习混合模型的 DDoS 攻击检测方法——DCNN-DSAE。与传统检测方法相比，该方法检测精度更高、误报率更低，且可以直接部署于 SDN 控制器中，具有很好的实用性。

3 特征提取及构建

在传统的机器学习分类方法中，输入特征是手动设计的，输入特征的好坏对模型的检测精度影响较大，好的特征的提取需要经过繁杂的运算和经验判决。而在深度学习中，模型可以自动地逐层提取多个不同层次的特征，并且将这些特征在不同层面组合起来产生输出。

由于深度学习模型可以自动地提取特征，因此直接提取交换机中流表的部分特征字段，作为模型第一层输入特征的一部分。自动获取的流表特征向量如表 1 所示。

表 1 自动获取的流表特征向量

特征名称	描述
durationSeconds	持续时间
packetCount	每条流中数据分组数量
byteCount	每条流中数据分组比特数
match.eth_type	匹配的以太网类型
match.eth_dst	匹配的以太网目的地址
match.eth_src	匹配的以太网源地址
match.in_port	匹配的入端口
priority	优先级
match.ipv4_src	匹配的源 IP
match.ipv4_dst	匹配的目的 IP
match.udp_dst	匹配的 UDP 目的端口
match.udp_dst	匹配的 UDP 源端口
match.tcp_dst	匹配的 TCP 目的端口
match.tcp_dst	匹配的 TCP 源端口
idleTimeoutSec	空闲超时
match.ip_proto	匹配的 IP 协议
durationNSeconds	持续时间
cookie	Cookie
tableId	表 ID
hardTimeoutSec	严格超时
actions.actions	动作

同时，为了提高模型检测精度、确保结果的可靠性，手动构建了最可能区分流类型的部分流表特征向量，并将其作为模型输入特征数据集的另一部分。表 2 是更详细的手动构建的流表特征向量。

表 2　　手动构建的流表特征向量

特征名称	描述
growth rate of single flow (grsf)	单流增长速率
growth rate of different port (grdp)	不同端口增长速率
average packets per flow (appf)	流表平均数据分组量
average bytes per flow (abpf)	流表平均比特数
average durations per flow (adpf)	流量平均持续时间

1) 单流增长速率（growth rate of single flow）

DDoS 攻击发生时，流表中单流和对流的增长速率常常成为检测攻击的重要特征。给定任意流 A 和流 B，当满足流 A 的源地址等于流 B 的目的地址、流 A 的目的地址等于流 B 的源地址、流 A 和流 B 具有相同的通信协议这 3 个条件时，流 A 和流 B 构成对流。

DDoS 常以 IP 欺骗的方式发起攻击，该特征增加了单流进入网络的数量，因为它们使用假 IP 地址发送数据分组，如式(1)所示。

$$grsf = \frac{\sum_{i=0}^{flow\_nums_T} nums\_single\_flows}{T} \tag{1}$$

其中，$flow\_nums_T$ 表示在时间周期 T 内采集的流表数量，T 表示采集周期，$nums\_single\_flows$ 表示每个周期 T 内采集到的流表中单流的数量。

2) 不同端口增长速率（growth rate of different port）

类似地，与 DDoS 攻击产生的 IP 欺骗相同，攻击者也可通过随机生成端口进行端口扫描攻击，如式(2)所示。

$$grdp = \frac{\sum_{i=0}^{flow\_nums_T} nums\_different\_port}{T} \tag{2}$$

其中，$nums\_different\_port$ 表示每个周期 T 内采集到的流表中不同端口号的流表数量。

3) 流表平均数据分组量（average packets per flow）

DDoS 攻击通过源 IP 欺骗，产生大量的、伪装不同 IP 的数据分组，且不同 IP 对应的数据分组数量较小。这种攻击方式使溯源任务非常困难，但是这也成为了区分正常分组和攻击分组的重要特征，因为正常分组中同一个 IP 对应的数据分组数量较大。鉴于这种特性，把流表平均数据分组量作为手动构建流表特征向量的重要特征之一，如式(3)所示。

$$appf = \frac{\sum_{i=0}^{flow\_nums_k} pcount_i}{flow\_nums_k} \tag{3}$$

其中，$flow\_nums_k$ 表示采集的所有流表中第 k 种流表的数量，$pcount_i$ 表示每条流表中的数据分组数量。

4) 流表平均比特数（average bytes per flow）

与定义流表平均数据分组量相似，DDoS 攻击的另一个特点是不同 IP 对应的数据分组的比特数较小。例如，在 TCP 泛洪攻击中，攻击者会发送大量的 120 B 的数据分组攻击受害者，这种特征也为 DDoS 攻击检测提供了依据，如式(4)所示。

$$abpf = \frac{\sum_{i=0}^{flow\_nums_k} bcount_i}{flow\_nums_k} \tag{4}$$

其中，$bcount_i$ 表示每条流表中的数据分组比特大小。

5) 流量平均持续时间（average durations per flow）

DDoS 泛洪攻击发生时，不同 IP 对应的持续时间较短，如式(5)所示。

$$adpf = \frac{\sum_{i=0}^{flow\_nums_k} durations_i}{flow\_nums_k} \tag{5}$$

其中，$durations_i$ 表示每条流表的持续时间。

综上所述，将部分流表统计信息直接作为深度学习的部分输入特征{M}，{M}包含表 1 中的全部特征向量，同时依据流表统计信息手动构建部分输入特征{N}，{N}包含表 2 中的特征向量。将直接获取的输入特征{M}和手动构建的输入特征{N}进行合并及维度重构，构成新的深度学习输入特征{M，N}。

4　基于深度学习混合模型的 DDoS 检测

4.1　攻击检测及防御总体架构

DDoS 攻击深度学习混合分类检测及防御总体架构如图 1 所示。本文提出的 DCNN-DSAE 混合模型通过分析流表特征来进行正确的分类。其中，DCNN-DSAE 混合模型的第一级采用 DCNN 模型，第二级采用 DSAE 模型。在检测过程中，DCNN 模型的输出端将流表特征分为攻击和正常 2 种类型。对于攻击类型，系统直接采用异常清洗的操作进行处理；而对于判断为正常类型的流

表特征，由于模型判断可能会存在误差，为了确保系统的安全性，将这些判断为正常类型的流表特征再转送到 DSAE 模型中进行进一步检测，最后，使用 softmax 分类器将其分类为正常类型或攻击类型。在模型的总输出端，当检测到异常时，控制器可通过 OpenFlow 协议修改流表等操作来有效缓解网络异常，并将新的安全策略下发到 OpenFlow 交换机中。

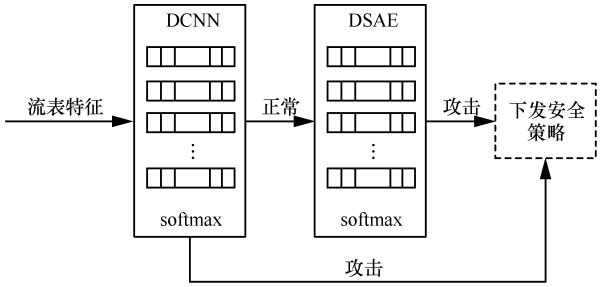


图 1 DDoS 攻击深度学习混合分类检测及防御总体架构

实际情况下，DCNN 模型在梯度下降计算时，可能会进入局部最优，或无限逼近全局最优，在模型的输出端不可能完全区分正常或 DDoS 攻击类型的流表特征，这意味着在 DCNN 模型的输出端，输出的正常类型的流表特征向量中可能会存在 DDoS 攻击类型的流表特征向量，即 DCNN 模型可能存在误判现象。

第一级 DCNN 模型采用监督学习方式。在训练 DCNN 模型的过程中，采用的数据集包括正常流量和 DDoS 攻击流量下的流表特征。正常流量下的流表特征标签为 0，DDoS 攻击流量下的流表特征标签为 1，模型的训练在有标签的监督下完成。第二级 DSAE 模型采用非监督的学习方式。在训练过程中，也采用和 DCNN 模型相同的数据集进行训练，但与之不同的是，DSAE 模型的训练不需要预先知道数据所对应的标签，在非监督的条件下完成流表特征的分类任务。相比于监督学习，非监督学习完全依赖于数据本身的特征，可以学习到更加抽象的特征。

DCNN-DSAE 混合模型汲取了 DCNN 和 DSAE 算法的优点，与传统的只采用卷积神经网络（CNN, convolutional neural network）[19]、自编码网络、自组织映射、支持向量机等异常检测方法相比，不仅可以提高分类检测的精度，还可以缩短分类检测的处理时间，从而有效地缓解了 SDN 中的 DDoS 攻击效应，避免资源耗尽。DCNN-DSAE 模型部署于 SDN 控制器中，采用流表特征作为模型的输入，输出为流表特征所对应的分类结果。

4.2 DCNN 模型

卷积神经网络是一种前馈人工神经网络，通常由一个或多个卷积层组成，然后由标准的多层神经网络中的一个或多个全连接层连接组成。CNN 的架构旨在利用输入数据的二维结构，这是通过本地连接和绑定权重实现的，然后通过池化，维持特征的平移、旋转、尺度的不变性。与具有相同层数的标准前馈神经网络相比，本文提出的 DCNN 模型具有更少的神经元和参数，更容易训练且检测精度更高。为了解决 DCNN 训练时间长的问题，本文的 DCNN 模型在训练过程中采用 GPU 加速和 2D 卷积高度优化，这样使模型的功能变得更加强大，能提取更高层次的流表特征，同时减少训练的时间，不会产生严重的过拟合问题。

通过实验对比，建立的 DCNN 网络模型如图 2 所示，其包含 3 个卷积层、2 个最大池化层以及 2 个全连接层。

输入的流表特征向量经第一层卷积层以提取更加抽象的高维度特征，然后对这些特征进行批标准化（batch normalization）处理，让模型学习到数据的分布特征；再经过第二层最大池化层后，保持

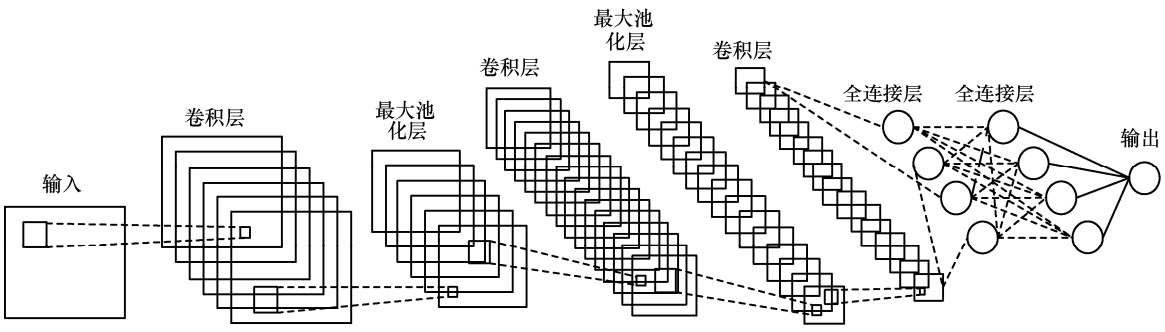


图 2 DCNN 模型结构

那些批标准化处理后的抽象高维特征的平移、旋转、尺度不变性，同时又能够减少模型的参数和计算量，防止出现过拟合现象，提高模型泛化能力；经过第三层卷积层，并将第三层卷积层抽象后的特征进行批标准化处理；再经过第四层最大池化层、第五层卷积层后，得到可准确表示流表特征的更高维度的特征向量；最后将这些高维特征向量输入全连接层，在输出端通过 softmax 分类器进行分类。

为防止模型训练时收敛速度较慢、出现梯度爆炸等现象，本文采用批标准化提升模型的容纳能力。如式(6)所示，每次按小批量训练模型，对每一批的数据做规范化处理，使输出的结果均值为 0、方差为 1。

$$\mu_B=\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i,\sigma_B^2=\frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{x}_i-\mu_B\right)^2 \tag{6}$$

批标准化在训练的过程中会带来梯度之间的相互竞争，这种梯度间竞争的好处是，模型在梯度下降减小误差更新权重和偏置的过程中避免陷入局部最优。设在前向传递的过程中，小批量的流表特征集合 B 中有 m 个流表特征向量，$B=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\cdots,\boldsymbol{x}_m\}$，对流表特征集合中 m 个流表特征向量按照式(7)进行归一化后得到归一化的流表特征集合 $\{\hat{\boldsymbol{x}}_1,\hat{\boldsymbol{x}}_2,\hat{\boldsymbol{x}}_3,\cdots,\hat{\boldsymbol{x}}_m\}$，最后，对归一化后的 m 个特征向量按照式(7)进行线性变换得到批标准化输出 $\{\boldsymbol{y}_1,\boldsymbol{y}_2,\boldsymbol{y}_3,\cdots,\boldsymbol{y}_m\}$。

$$\hat{\boldsymbol{x}}_i=\frac{\boldsymbol{x}_i-\mu_B}{\sqrt{\sigma_B^2+\epsilon}},\ \boldsymbol{y}_i=\gamma\hat{\boldsymbol{x}}_i+\beta\equiv BN_{\gamma,\beta}\left(\boldsymbol{x}_i\right) \tag{7}$$

其中，ε为常量，γ、β 为可学习的参数。

为进一步加快模型的训练速度，本文采用非饱和非线性激活函数修正线性单元 ReLU（式(8)）和 Softplus（式(9)）以代替传统的饱和非线性激活函数 Tanh 和 Sigmoid。

$$y=\max\left(0,x\right) \tag{8}$$

$$y=\ln\left(1+\mathrm{e}^x\right) \tag{9}$$

在当前模型深度和神经元个数及输入为大批量流量特征下，使用传统的饱和非线性激活函数会出现过拟合现象，在 DCNN 模型中进行梯度下降训练实验时，采用 ReLU 或 Softplus 这些非饱和非线性激活函数比采用 Sigmoid 或 Tanh 这些饱和非线性激活函数快好几倍。

为进一步克服模型的过拟合问题及提高模型的训练速度，本文使用了“dropout”[20]技术，即在神经网络传递过程中，将概率为 0.5 的隐藏神经元的输出设置为 0，因为这些神经元对前进传递没有贡献，所以将其丢弃，不再参与反向传播。因此，每次在批量数据送入模型训练时，神经网络都会随机地丢弃一定数量的神经元，使神经网络模型变得稀疏，且结构不同，但模型中所有这些体系结构共享权重。这种技术降低了模型对特定神经元结构的依赖性，因此，它迫使模型学习更多输入数据本身的特性。

4.3　DSAE 模型

自编码（AE, autoencoder）是一种前馈神经网络，它具有一个或多个隐藏层。当模型具有一个隐藏层时，隐藏层是输入特征向量的抽象表示，相当于主成分分析；当模型具有多个隐藏层时，在前向传递训练过程中，每 2 层之间通过受限玻尔兹曼机进行预训练，正向训练完成后，再通过误差反向传递调整权重和偏置，最小化输入和输出之间的差异。图 3 为自编码模型结构。

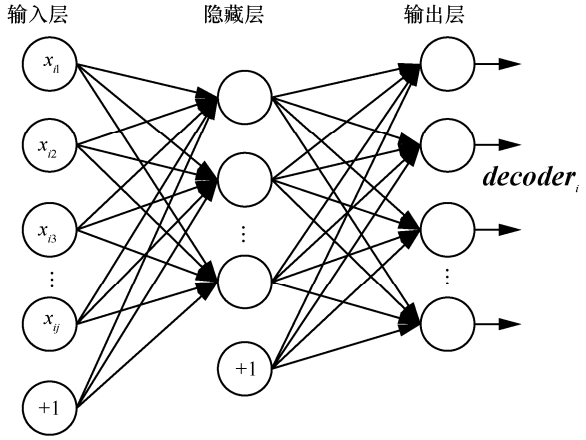


图 3　自编码模型结构

自编码模型由一个输入层、一个隐藏层和一个输出层组成。流表特征向量 $\boldsymbol{x}_i=\left[x_{i1},x_{i2},x_{i3},\cdots,x_{ij}\right]^{\mathrm{T}}$ 输入自编码模型的输入层，其中，i 表示第 i 个流表特征向量，j 表示每个流表特征向量中包含 j 个特征。隐藏层按照式(10)对输入的流表特征向量进行编码压缩。

$$\boldsymbol{encoder}_i=W_1\boldsymbol{x}_i+b_1 \tag{10}$$

其中，W_1 为连接输入层和隐藏层之间的权重，$\boldsymbol{x}_i$ 为输入的第 i 个流表特征向量，b_1 为隐藏层神经元的偏置。

编码完成后，基于隐藏层的输出结果，输出层根据式(11)进行解码重建，产生具有与输入层神经元相同尺寸的输出。

$$\boldsymbol{decoder}_i = f\left(W_2\left(\boldsymbol{encoder}_i\right)+b_2\right) \tag{11}$$

其中，f 为激活函数，W_2 为连接隐藏层和输出层之间的权重，$\boldsymbol{encoder}_i$ 为通过隐藏层编码压缩后的流表特征向量，b_2 为输出层神经元的偏置。

最后通过最小化损失函数（式(12)）达到训练自编码模型的目的。

$$loss=\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{decoder}_i\right)^2 \tag{12}$$

其中，n 为流表特征向量的个数，$\boldsymbol{x}_i$ 为输入的流表特征向量，$\boldsymbol{decoder}_i$ 为 $\boldsymbol{x}_i$ 经过自编码模型后输出的流表特征向量。

自编码模型的输入层和隐藏层神经元采用线性激活函数线性输出，输出层神经元采用 Sigmoid 函数（式(13)）非线性输出。

$$y=\frac{1}{1+\mathrm{e}^{-x}} \tag{13}$$

实验发现，当输出层神经元使用 Sigmoid 激活函数时，在检测精度上比使用 ReLU 或 Softplus 等激活函数效果更好，Sigmoid 函数在线性瓶颈问题上解决得更好，所得到的模型更易训练，模型对参数的变化也更具顽健性。

本文在构建模型时为了达到降维和提取抽象特征的目的，使用基于 AE 的 DSAE 模型。DSAE 模型通过逐层叠加自编码模型的输入层和隐藏层而构建，其中每个自编码模型采用一个隐藏层。DSAE 模型结构如图 4 所示。流表特征向量经过第一个自编码模型的学习后，在其隐藏层中得到压缩后的抽象化特征，将第一个自编码模型的隐藏层作为第二个自编码模型的输入层；经过第二个自编码模型的学习后，在其隐藏层中得到进一步压缩后的更加抽象化的特征，再将第二个自编码模型的隐藏层作为第三个自编码模型的输入层；依次叠加，最后通过 softmax 分类器将流表特征向量分为正常类型或 DDoS 攻击类型。

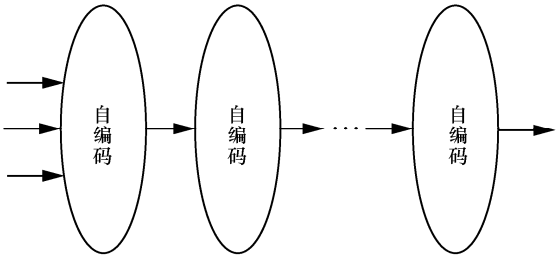


图 4 DSAE 模型结构

5 实验与结果分析

5.1 实验环境及评估指标

本文实验基于 Tensorflow 框架构建深度学习混合模型，硬件环境为 NVIDIA Tesla M40 类型的 GPU 服务器，软件环境为 Ubuntu16.04 操作系统。

实验前收集流表特征数据集，分别在正常流和 DDoS 攻击流下收集到自动获取和手动构建的流表特征，数据集如表 3 所示。流表特征数据集的总数为 15 万条，其中，训练集为 9 万条，测试集为 6 万条。

表 3 流表特征数据集

数据集	正常流特征/条	DDoS 流特征/条
训练集	43 000	47 000
测试集	28 000	32 000

实验通过准确度（accuracy）、精确度（precision）、召回率（recall）、F_1 分数（F_1 score）和混淆矩阵（confusion matrix）5 个评估指标来评估模型的检测性能。其中，TP（true positive）是实际类型为 DDoS 攻击的样本中被分类模型判断正确的样本数；TN（true negative）是实际类型为正常的样本中被分类模型判断正确的样本数；FN（false negative）是实际类型为 DDoS 攻击的样本被分类模型误判为正常类型的样本数；FP（false positive）是实际类型为正常的样本中被分类模型误判为 DDoS 攻击类型的样本数。

准确度（Acc）表示模型判断正确的数据分组的数量占数据分组总数量的百分比，即

$$Acc=\frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

精确度（P）表示模型判断为攻击类型的数据分组中，真正为攻击分组的数量所占的百分比，即

$$P=\frac{TP}{TP+FP} \tag{15}$$

召回率（R）表示模型判断为攻击类型的数据分组占所有攻击类型数据分组数量的百分比，即

$$R=\frac{TP}{TP+FN} \tag{16}$$

F_1 分数（F_1）表示精确度和召回率的调和平均值，能够更准确地评估模型性能。

$$F_1 = \frac{2PR}{P+R} \tag{17}$$

混淆矩阵（confusion matrix）主要用于模型分类的结果和数据与实际标签相匹配的程度。

5.2　DCNN 模型实验结果分析

在建立 DCNN 模型时，采用不同深度的卷积层都会对模型的检测精度产生较大影响，而且小批量训练的模型表现较好，模型训练时采用的 batch_size 大小默认为 50。实验过程中建立了 4 种不同深度的 DCNN 模型。4 种模型结构如表 4 所示。其中，C3P2F2 表示模型包含 3 个卷积层、2 个最大池化层和 2 个全连接层，1 表示模型存在该结构，0 表示模型不存在该结构。

本文对构建的表 4 中的 DCNN 模型进行实验，并通过定义的准确度、精确度、召回率、F_1 分数评估指标评估模型的性能。4 种不同深度的 DCNN 模型评估指标如表 5 所示。

采用3层卷积层的神经网络模型(C3P2F2、C3P3F2）效果明显优于采用 2 层卷积层的神经网络模型（C2P2F2、C2P2F4）。其中，C3P2F2 模型在精确度、召回率和 F_1 分数上的指标均高于 C3P3F2 模型，但在准确度上略低于 C3P3F3 模型。

为比较 C3P2F2 模型和 C3P3F2 模型的检测效果，通过混淆矩阵分析 4 种模型。4 种 DCNN 模型的混淆矩阵如图 5 所示。采用 2 层卷积层的神经网络的模型在判断正常分组和攻击分组的能力上均弱于采用 3 层卷积层的神经网络的模型。同时 C3P2F2 模型判断攻击分组的能力（0.98）强于 C3P3F2 模型（0.96）。考虑到 DDoS 攻击可能会危害到整个系统的安全性，采用的检测模型必须要对攻击有很强的敏感性，综合考虑，DCNN 模型采用 C3P2F2 模型最佳。

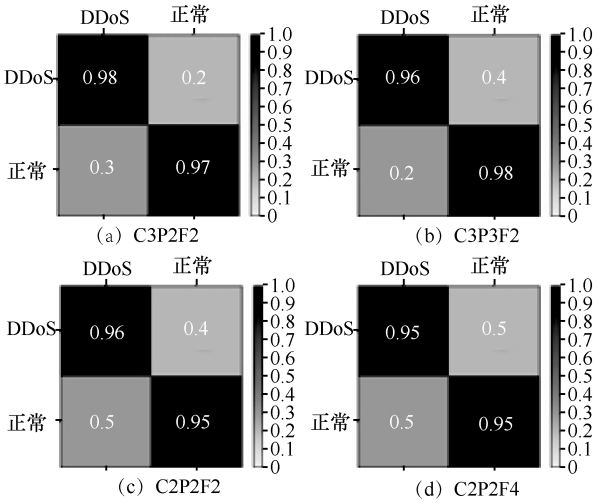


图 5　4 种 DCNN 模型的混淆矩阵

综上所述，C3P2F2 模型层数较深、神经元个数较多、训练模型采用的流表特征数据集较大，这些会导致模型在训练时收敛速度较慢。为加快训练的速度、提高模型的精度，对比进行批标准化处理和未标准化处理时模型的表现。实验结果如图 6 所示。

表 4　　4 种不同深度的 DCNN 模型的模型结构

模型名称	模型结构											
	卷积层 1	批标准化 1	池化层 1	卷积层 2	批标准化 2	池化层 2	卷积层 3	批标准化 3	池化层 3	全连接层（2）	全连接层（4）	激活函数
C3P2F2	1	1	1	1	1	1	1	0	0	1	0	ReLU, Softplus
C3P3F2	1	1	1	1	1	1	1	1	1	1	0	ReLU, Softplus
C2P2F2	1	1	1	1	1	1	0	0	0	1	0	ReLU, Softplus
C2P2F4	1	1	1	1	1	1	0	0	0	0	0	ReLU,Softplus

表 5　　4 种不同深度的 DCNN 模型的评估指标

模型名称	*Acc*	*P*	*R*	F_1
C3P2F2	97.73%	98.11%	97.93%	97.30%
C3P3F2	97.74%	97.17%	97.22%	97.02%
C2P2F2	96.43%	97.42%	97.54%	96.93%
C2P2F4	96.37%	97.07%	96.98%	97.01%

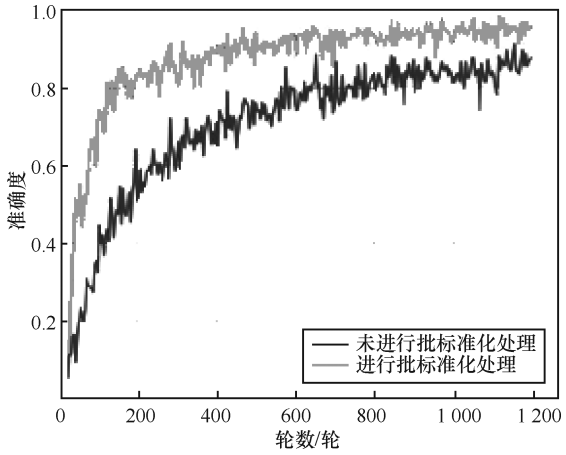


图 6 是否采用批标准化处理对模型准确度的影响

由图 6 可知，模型中加入批标准化处理后，加快了模型的训练速度，在训练到 180 轮时，准确度就已经达到 0.8，而没有进行规范化处理的模型，在 600 轮的训练后准确度才达到 0.8，并且加入批标准化后模型的精度表现更好。

同时，本文采用非饱和非线性激活函数修正线性单元 ReLU 和 Softplus 代替传统的饱和非线性激活函数 Tanh 和 Sigmoid。2 种激活函数的检测精度对比情况如图 7 所示。DCNN 模型在采用饱和非线性激活函数进行训练时，出现了较大的振荡，且检测精度低于采用非饱和非线性激活函数的训练结果。实验结果表明，DCNN 模型采用非饱和非线性激活函数对提高模型的检测精度更有效。采用 ReLU 等非饱和非线性激活函数，DCNN 模型的检测精度可以达到 0.98；而采用 Sigmoid 等饱和非线性函数，DCNN 模型出现了梯度消失的现象，甚至不能学习到有效的特征。

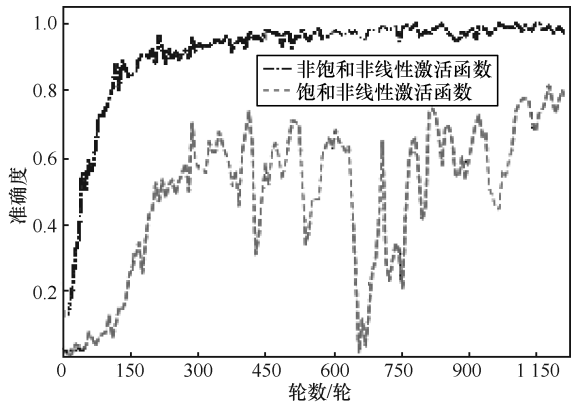


图 7 2 种激活函数的准确度对比情况

为克服训练过程中出现的过拟合问题，同时加快模型的训练速度，本文采取“dropout”技术有效地避免了模型训练过程中的过拟合问题[21]，实验结果如图 8 所示。模型在进行 1 200 轮训练的过程中，采用“dropout”训练时，梯度没有出现消失，误差一直在下降；而没有采用“dropout”时，模型的训练在进行到 600 轮时，就已出现梯度消失的现象，训练和测试结果如图 9 所示。采用“dropout”技术的模型在训练集合测试集上表现较好。

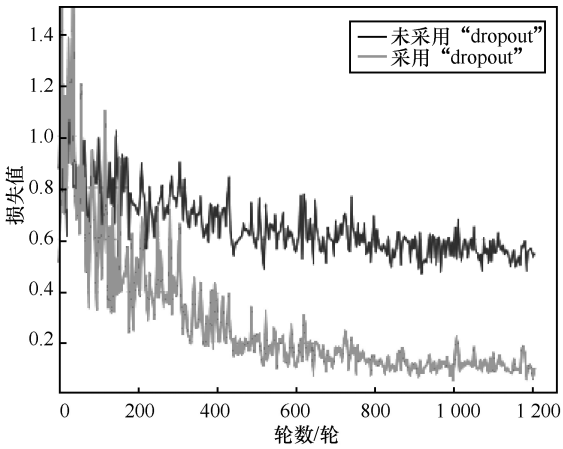


图 8 “dropout”技术对模型的影响对比

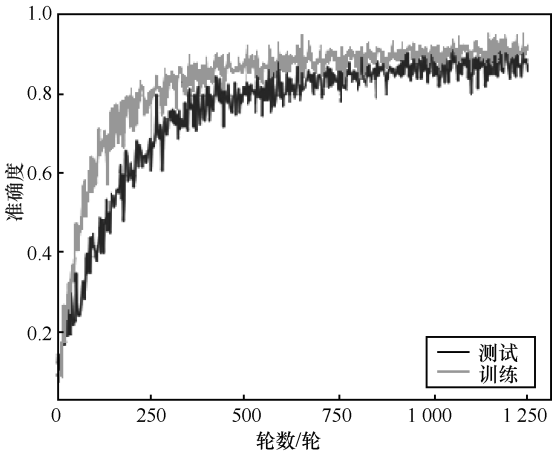


图 9 采用“dropout”的模型在训练集和测试集上的表现

综合所有结果可知，DCNN 模型采用 C3P2F2 模型为最佳，且模型中添加批标准化处理，激活函数采用非饱和非线性激活函数 ReLU 和 Softplus，在全连接层添加“dropout”防止过拟合。

实验将 C3P2F2 模型与传统机器学习方法中的 SVM 和深度学习方法中的深度神经网络（DNN, deep neural network）进行对比，并使用相同的数据集。3 种方法的精确度对比结果如图 10 所示。从图 10 可以看出，3 种模型在进行到 1 200 轮训练时都趋于收敛，且采用 C3P2F2 模型的检测精度明显高于 SVM 和 DNN 的方法。

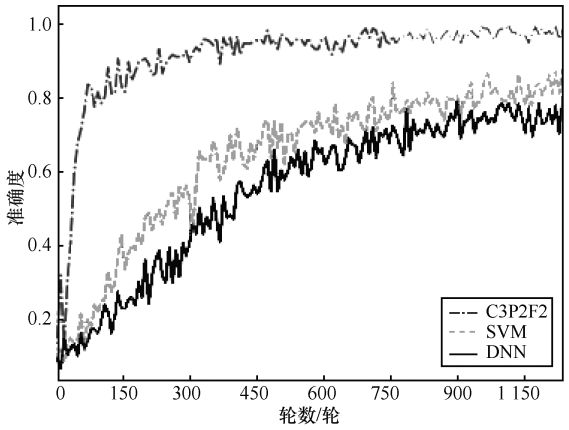


图 10　采用不同机器学习方法的检测精度对比

5.3　DSAE 模型实验结果分析

在构建 DSAE 模型时，采用自编码模型的个数会对模型的检测精度产生较大影响，实验过程中建立了 3 种不同深度的 DSAE 模型进行对比，为使模型表现得较好，采用小批量的训练方法，batch_size 大小默认为 50。具体所设计的 3 种不同深度的 DSAE 模型如表 6 所示。

SAE4 模型表示模型包含 4 个自编码模型，每个自编码模型都默认包含一个输入层、一个隐藏层和一个输出层。例如，在第一个自编码（128-64）中，128 表示输入、输出神经元个数为 128 个，隐藏层神经元为 64 个。表 6 中 1 表示模型存在该结构，0 表示模型不存在该结构。

表 6　3 种不同深度的 DSAE 模型的模型结构

模型名称	模型结构					
	自编码（128-64）	自编码（64-32）	自编码（32-16）	自编码（16-8）	自编码（8-4）	softmax
SAE3	1	1	1	0	0	1
SAE4	1	1	1	1	0	1
SAE5	1	1	1	1	1	1

表 7 为 3 种不同深度的 DSAE 模型评估指标。实验结果表明，采用 4 层自编码的 SAE4 模型在准确度、精确度、召回率和 F_1 分数上的指标均高于其他模型。

表 7　3 种不同深度的 DSAE 模型的评估指标

模型名称	*Acc*	*P*	*R*	F_1
SAE3	97.67%	97.41%	97.63%	97.30%
SAE4	98.11%	97.77%	97.74%	97.62%
SAE5	96.43%	97.72%	97.44%	96.95%

使用相同的数据集对 DCNN 的 C3P2F2 模型及 DSAE 的 SAE4 模型的实验结果进行对比。2 种模型的对比结果如图 11 所示。从图 11 可以看出，总

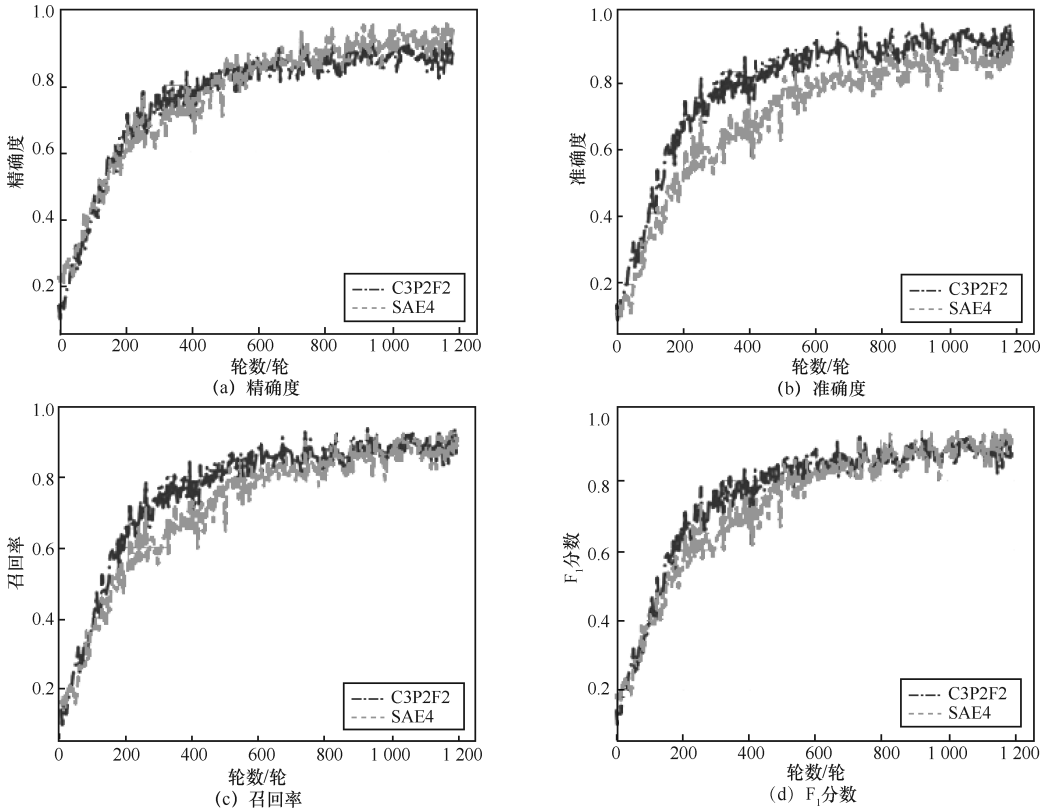


图 11　C3P2F2 模型和 SAE4 模型表现对比

体而言，C3P2F2 模型的综合评估指标略高于 SAE4 模型，这也是本文采用 C3P2F2 模型作为混合模型第一级的主要原因之一。

图 12 为 2 种模型的混淆矩阵对比。从实验数据可以看出，SAE4 模型判断攻击流的能力（0.985）强于 C3P2F2 模型（0.980）。考虑到第二级模型的主要功能是解决由于 DCNN 模型可能进入局部最优而导致的误判现象，区分被 DCNN 模型误判为正常流的攻击流，所采用的检测模型必须具备对攻击有很强的敏感性，综合考虑，第二级模型采用 SAE4 模型。

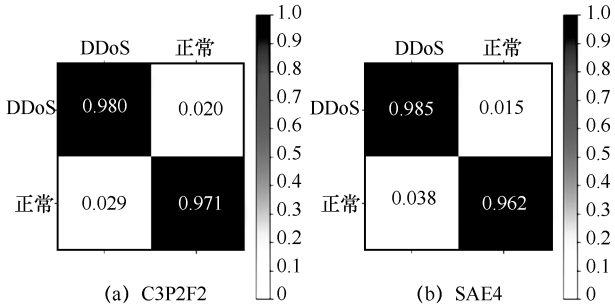


图 12 C3P2F2 模型和 SAE4 模型的混淆矩阵对比

5.4 DCNN-DSAE 模型总精度分析

将本文构建的 DCNN-DSAE 混合模型和传统的机器学习方法中的 SVM 及 DNN 进行对比分析，均使用相同的流表特征数据集作为模型的输入。实验检测结果如表 8 所示。

表 8 不同机器学习方法与 DCNN-DSAE 模型的评估指标

模型名称	*Acc*	*P*	*R*	F_1
SVM	96.72%	96.47%	95.84%	96.77%
DCNN-DSAE	98.53%	98.17%	97.94%	98.12%
C3P2F2	97.73%	98.11%	97.93%	97.30%
SAE4	97.44%	97.12%	97.87%	97.53%
DNN	94.47%	95.82%	93.74%	93.55%

从表 8 可以看出，DCNN-DSAE 模型在准确度、精确度、召回率和 F_1 分数上的指标均高于 SVM 和 DNN 模型，且 DCNN-DSAE 模型的准确度最高能达到 98.53%，高于只采用单级模型 C3P2F2（97.73%）和 SAE4（97.44%）的检测结果。从实验结果还可以看出，DCNN-DSAE 混合模型更优于传统的机器学习方法及单个 DCNN 或 DSAE 方法。

此外，输入的流表特征数量也是衡量 DDoS 攻击检测方法效果的重要参数，本文分别用 5、10、15、21、26、30 个流表特征对 DCNN-DSAE 模型进行测试。检测结果如表 9 所示。

表 9 不同输入特征数量下 DCNN-DSAE 模型的评估指标

特征数量/个	*Acc*	*P*	*R*	F_1
5	88.33%	89.15%	89.97%	90.07%
10	94.21%	94.19%	93.77%	94.04%
15	95.13%	95.20%	94.94%	94.85%
21	92.47%	92.57%	93.10%	92.20%
26	98.53%	98.17%	97.64%	98.12%
30	98.51%	98.20%	97.59%	98.11%

从表 9 可以看出，输入的流表特征数量对模型的检测结果存在重要的影响。当仅采用 5 个手动构建的流表特征（即流表特征数量为 5）作为 DCNN-DSAE 模型的输入特征时，DCNN-DSAE 模型的准确度只达到 88.33%；当同时采用 21 个自动获取的流表特征（即流表特征数量为 21）作为 DCNN-DSAE 模型的输入特征时，DCNN-DSAE 模型的准确度只达到 92.47%；当同时采用自动获取的流表特征和手动构建的流表特征作为 DCNN-DSAE 模型的输入特征时，DCNN-DSAE 模型的准确度得到有效提高；当采用 21 个自动获取的流表特征和 5 个手动构建的流表特征（即流表特征数量为 26）作为 DCNN-DSAE 模型的输入特征时，模型在准确度、精确度、召回率和 F_1 分数上的指标达到最高。

6 结束语

本文提出了一种基于深度学习混合模型的 DDoS 攻击检测方法——DCNN-DSAE，并重点分析了混合模型中 DCNN 模型和 DSAE 模型的构建过程和方法。通过实验，使用含攻击流和正常流的数据集进行模型测试，精确度可达到 98.53%，验证了此深度学习混合模型在实时网络环境中对 DDoS 攻击检测和防御的有效性。与传统的机器学习中的 SVM 方法和 DNN 方法相比，深度学习混合模型的方法检测精度更高、误报率更低，弥补了现有 DDoS 攻击检测方法的不足。由于该方法的检测输入特征为 SDN 交换机中的流表特征和自我构建的流表特征，属于轻量级的检测方法，可以直接部署于 SDN 控制器中。

参考文献：

[1] YAN Q, YU F R, GONG Q, et al. Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: a survey, some research issues, and challenges[J]. IEEE Communications Surveys & Tutorials, 2016, 18(1): 602-622.

[2] RADWARE. 2017-2018 global application & network security report[R]. 2018.

[3] AKAMAI. [State of the Internet]/security Q4 2017 executive summary[R]. 2017.

[4] VOELLMY A, WANG J. Scalable software defined network controllers[J]. ACM SIGCOMM Computer Communication Review, 2012, 42(4): 289-290.

[5] PENG T, LECKIE C, RAMAMOHANARAO K. Survey of network-based defense mechanisms countering the DoS and DDoS problems[J]. ACM Computing Surveys, 2007, 39(1):3.

[6] MIRKOVIC J, MARTIN J, REIHER P. A taxonomy of DDoS attacks and DDoS defense mechanisms[J]. ACM SIGCOMM Computer Communication Review, 2001, 34(2): 39-53.

[7] LI D, LI J, HUANG J, et al. Recent advances in deep learning for speech research at Microsoft[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013: 8604-8608.

[8] YU K. Large-scale deep learning at Baidu[C]//22nd ACM international conference on Information & Knowledge Management. 2013: 2211-2212.

[9] 杨余旺, 杨静宇, 孙亚民. 分布式拒绝服务攻击的实现机理及其防御研究[J]. 计算机工程与设计, 2004, 25(5): 657-660.
YANG Y W, YANG J Y, SUN Y M. Defense study and implementation mechanism of distributed denial of service attack[J]. Computer Engineering and Design, 2004, 25(5): 657-660.

[10] 孟江涛, 冯登国, 薛锐, 等. 分布式拒绝服务攻击的原理与防范[J]. 中国科学院大学学报, 2004, 21(1): 90-94.
MENG J T, FENG D G, XUE R, et al. Distributed denial of service attacks: principle and defense[J]. Journal of the Graduate School of the Chinese Academy of Sciences, 2004, 21(1): 90-94.

[11] GIL T M, POLETTO M. MULTOPS: a data-structure for bandwidth attack detection[C]//10th Usenix Security Symposium. 2001: 23-38.

[12] MOUSAVI S M, ST-HILAIRE M. Early detection of DDoS attacks against SDN controllers[C]//2015 International Conference on Computing, Networking and Communications (ICNC). 2015: 77-81.

[13] WANG R, JIA Z, JU L. An entropy-based distributed DDoS detection mechanism in software-defined networking[C]//2015 IEEE Trustcom/BigDataSE/ISPA. 2015: 310-317.

[14] JADIDI Z, MUTHUKKUMARASAMY V, SITHIRASENAN E, et al. Flow-based anomaly detection using neural network optimized with GSA algorithm[C]//2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops. 2013: 76-81.

[15] WINTER P, HERMANN E, ZEILINGER M. Inductive intrusion detection in flow-based network data using one-class support vector machines[C]//2011 4th IFIP International Conference on New Technologies, Mobility and Security. 2011: 1-5.

[16] TRUNG P V, HUONG T T, DANG V T, et al. A multi-criteria-based DDoS-attack prevention solution using software defined networking[C]//2015 International Conference on Advanced Technologies for Communications (ATC). 2015: 308-313.

[17] YUAN X Y, LI C H, LI X. DeepDefense: identifying DDoS attack via deep learning[C]//2017 IEEE International Conference on Smart Computing (SMARTCOMP). 2017: 1-8.

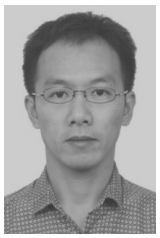
[18] 李传煌, 孙正君, 袁小雍, 等. 基于深度学习的实时 DDoS 攻击检测[J]. 电信科学, 2017, 33(7): 53-65.
LI C H, SUN Z J, YUAN X Y, et al. Real-time DDoS attack detection based on deep learning[J]. Telecommunications Science, 2017, 33(7): 53-65.

[19] LIU C, SUN W, CHAO W. Convolution neural network for relation extraction[C]//International Conference on Advanced Data Mining and Applications (ADMA 2013). 2013: 231-242.

[20] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.

[21] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[作者简介]



李传煌（1980−），男，江西九江人，博士，浙江工商大学副教授、硕士生导师，主要研究方向为软件定义网络、深度学习、开放可编程网络、系统性能预测和分析模型。

吴艳（1995−），女，安徽宣城人，浙江工商大学硕士生，主要研究方向为软件定义网络、深度学习。

钱正哲（1994−），男，浙江杭州人，浙江工商大学硕士生，主要研究方向为软件定义网络、深度学习。

孙正君（1993−），男，安徽滁州人，浙江工商大学硕士生，主要研究方向为软件定义网络、深度学习。

王伟明（1964−），男，浙江遂昌人，博士，浙江工商大学教授、硕士生导师，主要研究方向为新一代网络架构、开放可编程网络。