

基于深度残差胶囊网络与注意力机制的加密流量识别方法

史国振¹, 李昆阳¹, 刘瑶², 杨永健¹

(1. 北京电子科技学院, 北京 100070; 2. 西安电子科技大学计算机科学与技术学院, 陕西 西安 710071)

摘要: 随着用户安全意识的提高和加密技术的发展, 加密流量已经成为网络流量中的重要部分, 识别加密流量成为网络流量监管的重要部分。基于传统深度学习模型的加密流量识别方法存在效果差、模型训练时间长等问题。针对上述问题, 提出了一种基于深度残差胶囊网络模型 (DRCN, deep residual capsule network) 的加密流量识别方法。原始胶囊网络通过全连接形式堆叠导致模型耦合系数变小, 无法搭建深层网络模型。针对上述问题, DRCN 模型采用三维卷积算法 (3DCNN) 动态路由算法代替全连接动态路由算法, 减少了每个胶囊层之间传递的参数, 降低了运算复杂度, 进而构建深层胶囊网络, 提高识别的准确率和效率; 引入通道注意力机制为不同的特征赋予不同的权重, 减少无用特征对识别结果的影响, 进一步增强模型特征提取能力; 将残差网络引入胶囊网络层, 搭建残差胶囊网络模块缓解了深度胶囊网络的梯度消失问题。在数据预处理方面, 截取的数据包前 784byte, 将截取的字节转化成图像输入到 DRCN 模型中, 该方法避免了人工特征提取, 减少了加密流量识别的人工成本。在 ISCXVPN2016 数据集上的实验结果表明, 与效果最好的 BLSTM 模型相比, DRCN 模型的准确率提高了 5.54%, 模型的训练时间缩短了 232s。此外, 在小数据集上, DRCN 模型准确率达到 94.3%。上述实验结果证明, 所提出的识别方案具有较高的识别率、良好的性能和适用性。

关键词: 加密流量识别; 深度胶囊网络; 3D 卷积算法; 残差网络

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023007

Encrypted traffic identification method based on deep residual capsule network with attention mechanism

SHI Guozhen¹, LI Kunyang¹, LIU Yao², YANG Yongjian¹

1. Beijing Institute of Electronic Science and Technology, Beijing 100070, China

2. School of Computer Science and Technology, Xi'an University, Xi'an 710071, China

Abstract: With the improvement of users' security awareness and the development of encryption technology, encrypted traffic has become an important part of network traffic, and identifying encrypted traffic has become an important part of network traffic supervision. The encrypted traffic identification method based on the traditional

收稿日期: 2022-06-05; 修回日期: 2022-12-04

通信作者: 史国振, sgz1974@163.com

基金项目: 国家重点研发计划 (2017YFB0801803)

Foundation Item: The National Key R&D Program of China (2017YFB0801803)

引用格式: 史国振, 李昆阳, 刘瑶, 等. 基于深度残差胶囊网络与注意力机制的加密流量识别方法[J]. 网络与信息安全学报, 2023, 9(1): 32-41.

Citation Format: SHI Z G, LI K Y, LIU Y, et al. Encrypted traffic identification method based on deep residual capsule network with attention mechanism[J]. Chinese Journal of Network and Information Security, 2023, 9(1): 32-41.

deep learning model has problems such as poor effect and long model training time. To address these problems, the encrypted traffic identification method based on a deep residual capsule network (DRCN) was proposed. However, the original capsule network was stacked in the form of full connection, which lead to a small model coupling coefficient and it was impossible to build a deep network model. The DRCN model adopted the dynamic routing algorithm based on the three-dimensional convolutional algorithm (3DCNN) instead of the fully-connected dynamic routing algorithm, to reduce the parameters passed between each capsule layer, decrease the complexity of operations, and then build the deep capsule network to improve the accuracy and efficiency of recognition. The channel attention mechanism was introduced to assign different weights to different features, and then the influence of useless features on the recognition results was reduced. The introduction of the residual network into the capsule network layer and the construction of the residual capsule network module alleviated the gradient disappearance problem of the deep capsule network. In terms of data pre-processing, the first 784byte of the intercepted packets was converted into images as input of the DRCN model, to avoid manual feature extraction and reduce the labor cost of encrypted traffic recognition. The experimental results on the ISCXVPN2016 dataset show that the accuracy of the DRCN model is improved by 5.54% and the training time of the model is reduced by 232s compared with the BLSTM model with the best performance. In addition, the accuracy of the DRCN model reaches 94.3% on the small dataset. The above experimental results prove that the proposed recognition scheme has high recognition rate, good performance and applicability.

Keywords: encrypted traffic identification, deep capsule network, 3D convolutional algorithm, residual network

0 引言

随着用户隐私意识的提高以及网络流量加密技术的快速发展,加密流量已经成为网络流量的重要组成部分。加密流量爆炸式增长给网络空间安全带来了巨大隐患,尤其是许多恶意软件通过加密技术的保护绕过防火墙实现了对用户的攻击。加密流量识别技术^[1]是实现网络态势感知和网络安全监管的关键技术。加密流量识别已经成为一个热门研究领域,在学术界和工业界引起了广泛关注。

传统的流量识别方法可以分为3种:基于端口的方法^[2]、基于深度包解析(DPI, deep packet inspection)的方法^[3]、基于机器学习的方法^[4-6]。其中,基于端口的方法准确率较低,目前已经被淘汰;基于DPI的方法无法实现对加密流量的识别;基于机器学习的方法准确率很大程度上依靠人工提取的特征,但是如何构造一组优良的特征集仍是研究重点。因此上述3种方法很难应用到加密流量识别中。

由于深度学习不需要进行特征提取的特性,其在图像识别、语音处理和自然语言处理等领域均得到了广泛的应用并且取得了优良的结果,越来越多的研究人员尝试将深度学习方法应用到加密流量识别领域。文献[7]提出了一种基于一维卷

积神经网络(CNN, convolutional neural network)的端到端的加密流量分类方法,该方法将数据的预处理、加密流量特征的提取和模型的学习等整合到一个方法中,实现了对加密流量的端到端分类,但是由于所用模型结构简单,识别的准确率较低。文献[8]利用数据包到达时间和数据包大小组成二维图像,并使用卷积神经网络对这些图片组成的数据集进行分类,对于VPN流量取得了较好的识别结果,但是对于Tor流量的识别效果较差。文献[9]对循环神经网络(RNN, recurrent neural network)进行改进,提出了一种树结构循环神经网络,利用树形结构将大类划分成小类,然后对每个小类都设置一个特定的分类器,实现了分类性能的互补,但是缺少在不平衡数据集下的实验,无法验证模型的鲁棒性。文献[10]提出了基于循环神经网络的扩展字节神经网络,这种网络模型除了使用数据包的有效载荷信息外还充分利用了侧信道特征对加密流量进行分类。文献[11]提出了一种基于长短期记忆网络(LSTM, long short term memory network)和注意力机制的加密流量识别方法,通过引入注意力机制,使得模型更关注重要的数据包流。文献[12]提出了一种基于CNN+RNN的网络模型,其中CNN用于学习数据包内部的空间特征,RNN用于学习数据包之

间的时间特征,更为全面地利用了加密流量的特征,但是由于RNN的使用,模型的并行度较低,效率较差。通常来说,卷积神经网络能够较好地学习到流量数据的空间特征,难以学习数据包与数据包之间的时间特征。循环神经网络常用来处理序列数据,对于单个数据包中的关键信息学习能力较差。目前来说,使用CNN+RNN的方法能够较全面地利用加密流量的特征,但是由于RNN的并行性差,模型训练效率低,实时性差。

针对当前深度学习模型进行加密流量分类任务存在的缺点,本文采用胶囊网络^[13]作为基础的网络模型。胶囊网络使用的胶囊是由一组神经元构成的向量,不仅能够学习到单个数据包中的空间特征信息,而且能够捕捉到不同数据包之间的时间特征信息。同时胶囊网络在低层胶囊和高层胶囊之间采用动态路由的方式进行特征传递,有更强的细节特征提取能力。为了更好地学习加密流量的特征,提高模型识别的准确率,需要搭建深层胶囊网络。最近的研究表明^[14-15],胶囊网络中动态路由算法的计算开销非常大,多个胶囊层会导致模型的训练时间呈几何式增长,通过简单胶囊层堆叠来实现深度胶囊网络的方法是无效的。针对这个问题,本文将3D卷积算法引入动态路由机制,以卷积连接取代全连接减少了模型需要的参数,降低了计算开销和训练时间。此外,当胶囊层过多时,在梯度反向传播的过程中,靠近输入的神经元的梯度远远小于输出端的神经元,导致梯度消失。本文提出在深度胶囊网络中引入残差网络^[16],构成残差胶囊层,解决梯度消失问题,提高学习的准确率。同时,为了在高维度的加密流量数据中加强模型对于特征的提取能力,引入了注意力机制,进一步提升了模型对于加密流量识别的准确率。

本文的主要贡献如下。

1) 将通道注意力机制和胶囊网络相结合,全面利用了加密流量的空间特征和时间特征,提高了加密流量识别的准确率。

2) 将3D卷积算法和残差网络引入胶囊网络,搭建深度残差胶囊网络用于加密流量识别,提高了加密流量识别的效率,增强了识别的实时性,并进一步提高了识别的准确率。

1 相关理论

1.1 通道注意力机制

注意力机制能够从大量的信息中将重要的信息赋予更高的权重,从而实现减少无用信息和噪声对于结果的干扰,以提高模型分类的准确率。传统的注意力机制对于不同位置不同子空间的特征信息采用取平均值的方法,会导致某些重要特征丢失。而对于加密流量识别领域,数据包不同的特征对于识别效果起到了不同的作用,因此传统的注意力机制难以满足加密流量识别的需求。文献[17]提出了通道注意力机制,该机制能够选择性地增强携带更多信息的特征,并有效抑制无效特征,使得后续处理可以充分利用流量特征。

1.2 胶囊网络

胶囊网络这个概念是由Hinton在2017年提出来的,并在minist数据集上取得了最好的成绩。胶囊是由多个神经元组成的集合,一个神经元可以被看成标量,那么由多个神经元组成的胶囊就是向量。每个神经元可以分别表示图像中不同的属性,如方向、形状、颜色等信息,因此用胶囊作为网络的基本单元能够学习并传递更多的信息。在胶囊网络中,低层向高层的信息传递是通过动态路由的方法实现的,连接权值会在学习中发生变化,由此引起节点连接程度的变化。总体来说,相较传统的神经网络,胶囊网络具有更强的特征提取能力,更适用于网络流量分类这项任务。

2 基于深度残差胶囊网络的加密流量识别模型

VPN流量是最常见且应用广泛的加密流量,本文选取VPN流量作为加密流量的代表,应用基于DRCN模型的加密流量识别方法对其进行识别。本文采用端到端的模型架构,包括数据预处理、通道注意力和深度胶囊网络3个模块,整体架构如图1所示。

1) 数据预处理模块:根据五元组对加密流量数据包进行划分、清洗和可视化,将原始数据转换成深度学习模型能够识别的图片形式。

2) 通道注意力模块:运用通道注意力机制,对加密流量识别中关键的特征赋予更高的权重,

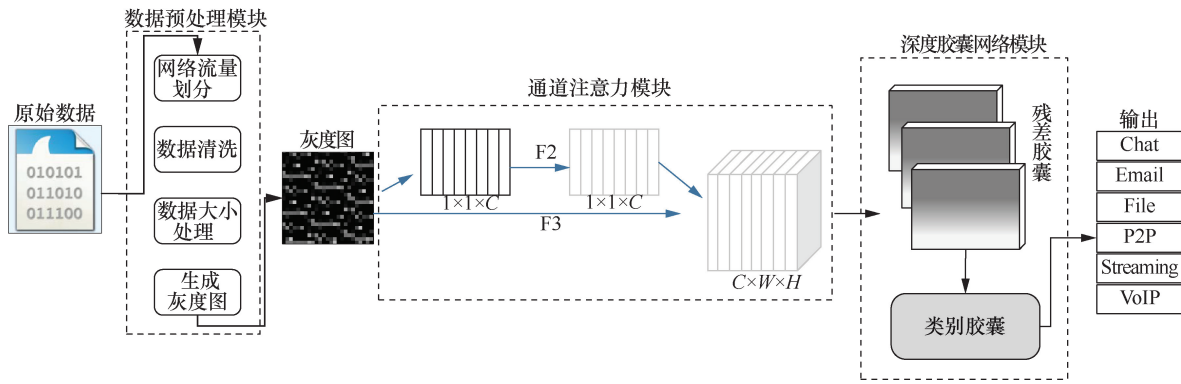


图 1 整体架构
Figure 1 Overall architecture

实现数据特征强化。

3) 深度胶囊网络模块：使用深度胶囊网络模型对处理后的加密流量数据进行识别和分类。

2.1 数据预处理模块

数据预处理是加密流量识别基础，直接关系到识别的准确率和效率。本文提出的数据预处理方法，不用进行人工的特征选择，只截取了数据包前 784 byte，实现将 pcap 格式文件转化成了 28×28 像素的灰度图片。

1) 流量划分：具有相同五元组^[18]（源 IP 地址、目的 IP 地址、源端口号、目的端口号和网络协议）的数据包属于同一种应用产生的网络流量，称为一个网络流。本文以网络流为单位对原始的流量文件进行划分。

2) 数据清洗：由于在进行加密流量采集时，使用的总是固定的某几台机器，因此 IP 地址和 MAC 地址是相对固定的。这两组地址不仅对加密流量识别任务起不到正面作用，还会引入负增益，降低模型的效果。因此，本文对所有数据包的 IP 地址和 MAC 地址进行了随机化处理。另外，在数据采集过程中，受到设备和环境的干扰，会有一些数据包出现损坏、丢失或者载荷为空的情况。在这一步骤中，要将这类数据包清理掉。

3) 数据大小处理：对于胶囊网络来说，输入长宽均为 28 像素的图片，且将数据包进行分类的有效信息都保存在数据包头部^[19]，因此本文选取每个数据包前 784 byte 数据。如果文件大小大于 784 byte，则截取前 784 byte；如果文件大小小于 784 byte，则在文件末尾，添加 0 补充文件至 784 byte。然后对字节数据进行归一化，得到 784 个 0 到 1

之间的数值。

4) 用黑色表示 0，白色表示 1，中间数值用灰度代替，形成灰度图片。

2.2 注意力机制模块

对于加密流量识别来说，并不是所有的特征都能够对识别产生相同的作用。为了将重要特征赋予更高的权重，本文引入通道注意力机制^[17]，模块示意如图 2 所示，以实现重点关注数据包长度、数据包持续时间、端口号等重要特征。

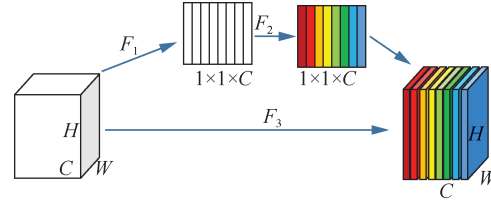


图 2 通道注意力机制模块示意
Figure 2 Schematic of channel attention mechanism

通道注意模块输入的维度为 $C \times H \times W$ ，通道数为 C ，其实现过程如下。

(1) 压缩过程

该模块在开始时利用压缩操作对每个通道上的信息点进行平均，计算平均后的数值，即对输入的每一个通道进行一次全局的均值池化过程，然后得到维度为 $1 \times 1 \times C$ 的压缩数据 z 。压缩数据 z 在第 c 个通道上的值 z_c 的计算过程如下。

$$z_c = F_1(u_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j) \quad (1)$$

其中， u_c 表示 U 在第 c 个通道上的输出特征矩阵。

(2) 激发过程

通过全连接层和激活函数将压缩数据 z 映射

到较小的 $1 \times 1 \times (C/r)$ 维度, 其中 r 为压缩比例。再通过上述的全连接层和激活函数恢复到初始的 $1 \times 1 \times C$ 的维度, 进而计算得到校准向量 $\mathbf{sc}$ 。校准向量的计算如式(2)所示。

$$\mathbf{sc} = F_2 = f_2(w_2 \times f_1(w_1 \times z)) \quad (2)$$

通道生成的校准向量 $\mathbf{sc}$, 用来表示通道间的相关性显示建模。其中, f_1 表示线性整流激活函数, f_2 表示 Sigmoid 激活函数。

(3) 校准过程

将校准向量 $\mathbf{sc}$ 和对应的通道相乘, 即可得到特征矩阵的重新定表, 从而获得输出特征 $\tilde{\mathbf{X}}_c$ 。然后利用 $\mathbf{sc}_c$ 表示校准向量的第 c 维度上的值, $\mathbf{u}_c$ 表示 U 在第 c 个通道上的输出特征矩阵, 则输出特征矩阵 $\tilde{\mathbf{X}}$ 的第 c 个通道特征矩阵的计算过程如下。

$$\tilde{\mathbf{X}}_c = F_3 = \mathbf{sc}_c \times \mathbf{u}_c \quad (3)$$

综上, 通道注意力模块的输入为 U , 输出为 $\tilde{\mathbf{X}}$ 。

2.3 深度胶囊网络模块

对于加密流量识别任务来说, 面临的一个难题是加密流量数据包特征维度较高, 简单的浅层胶囊网络难以有效捕捉到全部的特征信息。为了更好地利用加密流量数据包的特征信息, 需要构建深度胶囊网络。原始的胶囊网络是通过全连接的形式进行堆叠的, 但将全连接的胶囊层堆叠在一起, 会导致胶囊层之间的耦合系数变小, 抑制梯度传播和学习效果, 同时模型的训练时间会呈几何式增长。为了减少模型训练时间, 本文将 3D 卷积算法^[20]引入动态路由的过程, 用卷积的动态路由代替全连接的动态路由来实现深度胶囊网络。

设每一层胶囊为 $\Phi^l \in R^{(w^l, w^l, c^l, n^l)}$, 其中 w^l 表示特征图的宽度和高度, c^l 表示 3D 胶囊的数量, n^l 表示胶囊的大小, l 表示胶囊的层数。本文中两个胶囊层之间的动态路由算法通过 3D 卷积实现。首先, 将胶囊 Φ^l 表示成一个新的单通道张量 $\tilde{\Phi}^l \in R^{(w^l, w^l, c^l \times n^l, 1)}$, 并与 $c^l \times n^l$ 这个卷积核进行卷积。设 V 表示 3D 卷积的结果, 同时为了保证 V 的大小与输入的单通道张量 $\tilde{\Phi}^l$ 的通道数一致, 使用 $(1, 1, n^l)$ 作为 3D 卷积的步长。 V 中的每个元素 $v_{i,j,k,m}$ 可以通过式(4)计算得到。

$$v_{i,j,k,m} = \sum_p \sum_q \sum_r \tilde{\Phi}^l(i-p, j-q, k-r) \Psi_t^l(p, q, r) \quad (4)$$

其中, i, j, k 和 p, q, r 分别表示特征图和 3D 卷积核的空间位置, m 表示 l 层中 3D 胶囊的张量, Ψ_t^l 表示 l 层第 t 个卷积核。

通过卷积运算的特征图具有局部特征, 相邻的胶囊之间会携带相似的信息, 从而可以通过一组胶囊 s 由层 l 向高层 $l+1$ 传递信息。与传统的只能通过全连接形式在两层之间传递信息的动态路由算法相比, 基于 3D 卷积的动态路由算法 (如算法 1 所示) 只通过一组胶囊 s 即可传递信息, 减少了动态路由算法的参数和时间复杂度。

算法 1 3D 卷积动态路由算法

- 1) Require: $\Phi^l \in R^{(w^l, w^l, c^l, n^l)}$, r and c^l, n^l
- 2) $\tilde{\Phi}^l = \text{Reshape}(\Phi^l), \tilde{\Phi}^l \in R^{(w^l, w^l, c^l \times n^l, 1)}$
- 3) $V = \text{Conv3D}(\tilde{\Phi}^l)$
- 4) $\tilde{V} = \text{Reshape}(V)$
- 5) $B = 0, B \in R^{(w^{l+1}, w^{l+1}, c^{l+1}, c^{l+1})}$
- 6) 令 $p \in w^{l+1}, q \in w^{l+1}, r \in c^{l+1}, s \in c$
- 7) for i iterations do
- 8) for all $p, q, r, k_{pqrs} \leftarrow \text{softmax_3D}(b_{pqrs})$
- 9) for all $s, S_{pqr} \leftarrow \sum_s k_{pqrs} \cdot \tilde{V}_{pqrs}$
- 10) for all $s, S_{pqr} \leftarrow \leftarrow \text{softmax_3D}(S_{pqr})$
- 11) for all $s, b_{pqrs} \leftarrow b_{pqrs} + \tilde{S}_{pqr} \cdot \tilde{V}_{pqrs}$
- 12) 返回 $\Phi^{l+1} = \tilde{S}$

由于高层 $l+1$ 的每个胶囊均可以从低层获 l 得 c^l 个对应的预测结果, 每个预测结果乘以对应的耦合系数后产生唯一的预测值 S_{pqr} 。使用非线性归一化函数, 将向量的模长压缩在 $[0, 1]$ 范围内。

$$S_{pqr} = \sum_s k_{pqrs} \tilde{V}_{pqrs} \quad (5)$$

$$\hat{s}_{pqr} = \text{squash}_{3D(S_{pqr})} = \frac{S_{pqr}^2}{1 + S_{pqr}^2} \cdot \frac{S_{pqr}}{S_{pqr}} \quad (6)$$

从图 3 可以看出, 在向量模长接近 0 时, 压缩函数得到的结果较小, 在全局压缩的过程中, 会丢失某些有用的信息。

基于这个问题, 本文尝试对压缩函数进行改进, 分别尝试了 3 种不同的压缩函数。

$$F1 = \text{squash}_{3D1}(S_{pqr}) = \frac{S_{pqr}^2}{0.1 + S_{pqr}^2} \cdot \frac{S_{pqr}}{S_{pqr}} \quad (7)$$

$$F2 = \text{squash_}3D_2(S_{pqr}) = (1 - e^{-S_{pqr}}) \cdot \frac{S_{pqr}}{S_{pqr}} \quad (8)$$

$$F3 = \text{squash_}3D_3(S_{pqr}) = \frac{S_{pqr}}{0.1 + S_{pqr}} \cdot \frac{S_{pqr}}{S_{pqr}} \quad (9)$$

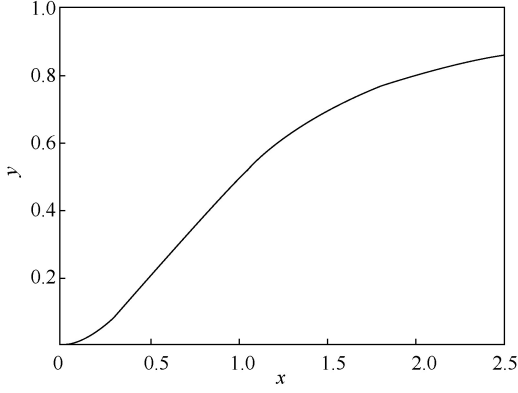


图3 压缩函数图像
Figure 3 Image of squash function

实现深度胶囊网络面临的另一个问题是胶囊层之间的耦合系数变小，抑制梯度传播和学习效果，本文将残差结构^[21]引入胶囊网络，构成残差胶囊层。

深度胶囊网络结构如图4所示，由1个卷积层、3个残差胶囊网络层和1个类别胶囊网络层组成。在卷积层中，采用了128个3×3的卷积核进行采样；在残差胶囊网络层中，使用了3个3D胶囊单元；类别胶囊通过判断模长最大的胶囊向量，输出识别结果。

每一个残差胶囊网络层由3个3D胶囊单元组成，其中主要通道上包含2个胶囊单元，残差通道上包含1个胶囊单元。由于胶囊是用向量表示的，因此两个通道相加时没有使用通道合并，而是将两个通道的输出矩阵进行对应元素数值相加，这种方式减少了计算过程中的误差，从而缓解了深度网络的梯度消失。同时两个相邻胶囊之间使用3D卷积动态路由算法取代了全连接路由算法。全连接路由算法要求在低层胶囊和高层胶囊之间传递特征信息时，所有的胶囊都要建立连接，因此每一层传递都需要保存大量的参数。而3D卷积产生的特征胶囊具有局部特征，在低层只需要选取一组不相邻的胶囊向高层胶囊传递信息即可，很大程度减少了需要的参数，进而降低了运算的复杂度。

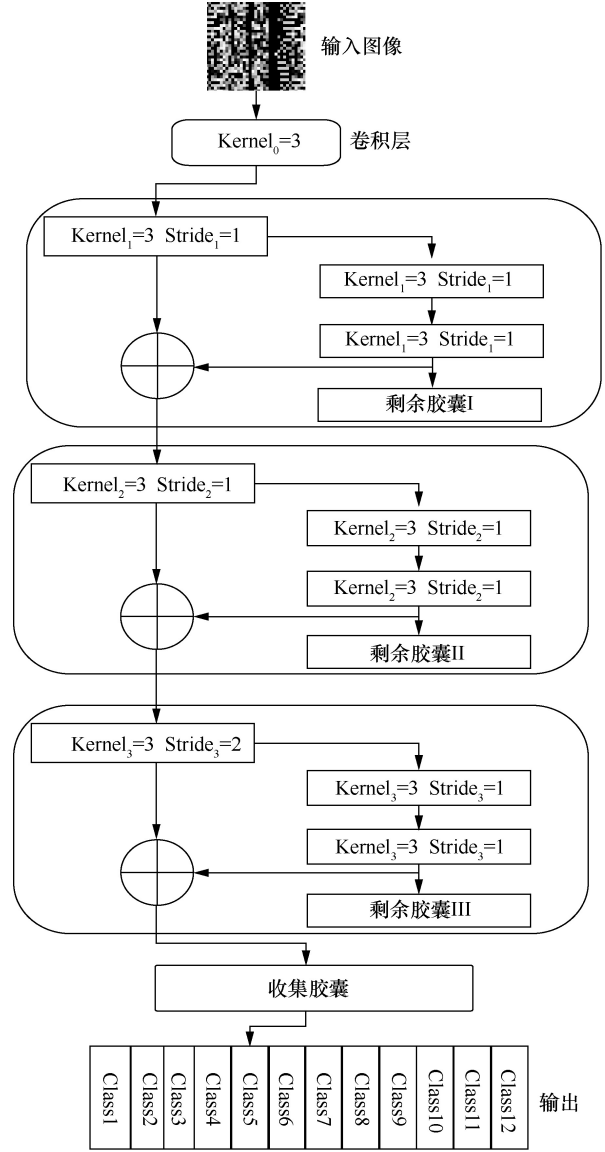


图4 深度胶囊网络结构
Figure 4 Deep capsule network

本文采用边缘损失函数作为本文的损失函数。

$$L_r = \lambda(1 - T_c) \max(0, v_c - m^-)^2 \quad (10)$$

$$L_l = T_c (0, m^+ - v_c)^2 \quad (11)$$

$$L_{\text{total}} = \sum_c (L_r + L_l) + ax - x^2 \quad (12)$$

其中， T_c 表示第 c 类结果出现的情况，出现时 T_c 取1，否则取0； v_c 表示第 c 类胶囊的模长； m^+ 和 m^- 分别表示阈值，一般情况下， m^+ 取0.9， m^- 取0.1，分别表示惩罚预测存在而实际不存在的错

误情况和用于惩罚实际存在而预测不存在的错误情况； λ 表示 m^+ 和 m^- 的权重，一般取 0.5。

3 实验与结果分析

3.1 实验环境

本文的实验环境为 Windows 10 操作系统，python 3.6 开发系统，CPU 为 AMD Ryzen 7 5800H，内存为 16 GB，显卡为 RTX3060，基于 pytorch1.10 完成模型构建。

深度胶囊网络模型的参数设置如下：训练的迭代次数为 50 次，采用 Adam 优化算法，初始学习率为 0.001。初始的卷积层中使用 128 个 3×3 的卷积核，进行特征提取。胶囊层包括残差胶囊层和类别胶囊层，胶囊的维度依次为 2、4、8。

为防止实验偶然性，实验结果采用五折交叉验证，训练集和测试集所占比例分别为 75% 和 25%。

3.2 实验数据

本文采用 University Of New Brunswick 公开的 ISCXVPN2016 数据集^[22]验证上述方法的有效性。该数据集是通过 Wireshark 和 Tcpdump 捕获现实中真实网络流量构成的，包括 7 种使用 VPN 加密的网络流量。由于该数据集 150 个 pcap 格式的原始文件是按照特定应用进行分类的，特别是其中 hangoutVoIP 这组数据，本文无法判断是属于 browsing 类型还是属于 VoIP 类型，因此在训练中将这组数据进行了剔除。经过筛选后，本文保留了 6 组常规的网络流量和 6 组使用 VPN 加密的网络流量。

从图 5 可以看出，不同流量数据之间的数量以及同种流量是否加密的数量相差很大，在这种不平衡数据集下进行实验可以进一步验证模型的鲁棒性。为了验证模型在小数据集情况下的效果，本文对每种流量类型随机抽取 1 000 条数据，构成了加密流量小数据集。

图 6 为小数据集的流量类别及数据量，每种流量类型均为 1 000 条。

3.3 评价指标

实验采用准确率、精确率和召回率作为实验结果的评价指标，其中，真阴性 TN 表示数据为正常且预测为正常的数量，真阳性 TP 表示数据为异常且预测为异常的数量，假阴性 FN 表示数据为异常但预测为正常的数量，假阳性 FP 表示数据为正常但预测为异常的数量。

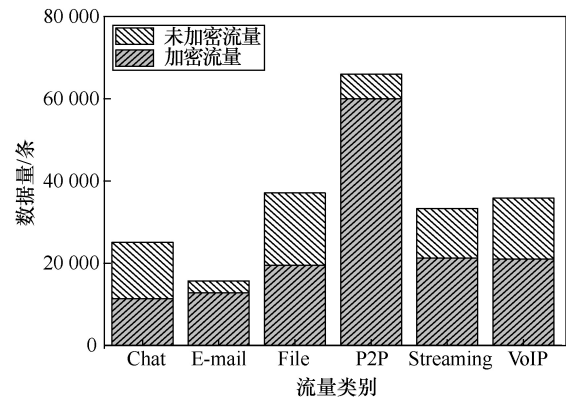


图 5 流量类别及数据量

Figure 5 Traffic categories and data volumes

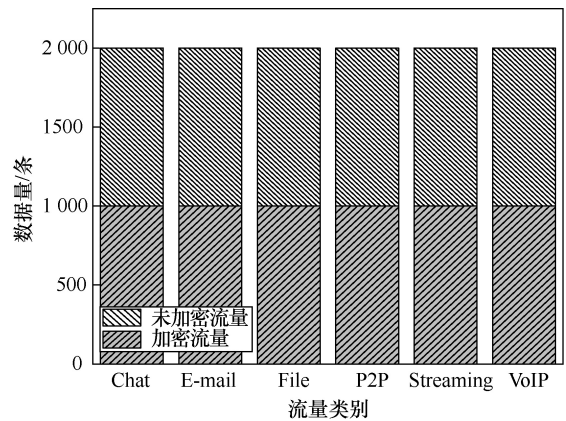


图 6 小数据集的流量类别及数据量

Figure 6 Traffic categories and data volumes for small dataset

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

Accuracy 表示准确率，该数值越大，表示模型的性能越好。

$$\text{Precision} = \frac{TP}{TP+FP} \quad (14)$$

Precision 表示精确度，该数值越高表示模型的误报率越低。

$$\text{Recall} = \frac{TP}{TP+FN} \quad (15)$$

Recall 表示召回率，该数值越高，表示模型的误报率越低。

3.4 结果分析

3.4.1 压缩函数选择与分析

不同的压缩函数对应的性能指标不同，为了提高模型的识别效果，本文选择 3 种不同的压缩

函数和原始的压缩函数进行比较实验,进而确定最优的压缩函数。

不同压缩函数压缩率如图7所示,从中可以看出,原始胶囊网络的压缩函数在自变量全局都实现了对模长的压缩,而本文使用的3种压缩函数,在自变量为0附近时,起到了放大模长的作用,放大了各个类别之间的间距,进而提高了模型的识别效果。

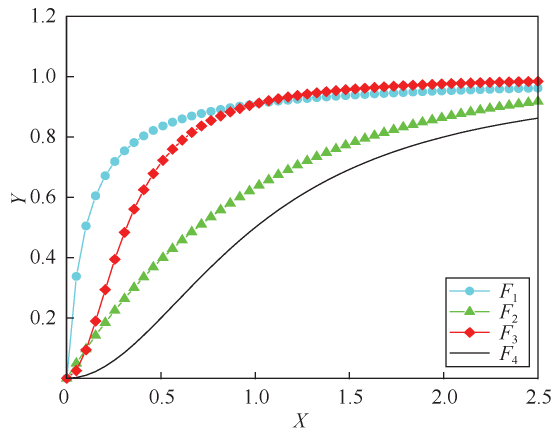


图7 不同压缩函数压缩率

Figure 7 Compression ratio of different squash functions

不同压缩函数准确率曲线如图8所示,从中可以看出,随着训练次数的增加,不同压缩函数情况下模型的准确率都在逐步提高,但原始压缩函数的识别准确率最低,且收敛速度最慢,而压缩函数F1在50轮左右即收敛。因此本文最终采用压缩函数F1作为DRCN模型的压缩函数。

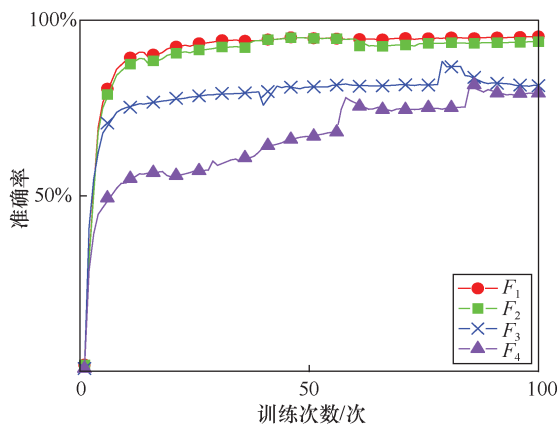


图8 不同压缩函数准确率曲线

Figure 8 Accuracy curves of different aquash functions

3.4.2 原始数据集和小数据集准确率对比

图9展示了基于深度胶囊网络对在不平衡数据集下加密流量识别结果的混淆矩阵,可以发现该模型对不同类型的流量数据都能够取得较好的识别结果。每一类流量数据的识别率都在90%以上,且对Novpn_Video流量数据的识别效果达到了100%。从图9能够发现本文模型存在以下问题:①对于Audio流量数据是否使用VPN加密的识别效果不理想,将3.37%没有使用VPN加密的chat数据识别成了使用VPN加密的流量数据,同时将3.06%使用VPN加密的chat流量识别成了没有使用VPN加密的流量;②由于Novpn_Email样本数据较少(只有2000条),对该种流量识别的效果是最差的,准确率只有93.91%。在后续的工作中,可以通过采集NovpnE-mail数据完善数据集,让模型在数据平衡情况下进行学习。

NovpnChat	0.9573	0.0090	0	0	0	0	0.0337	0	0	0	0	0	0	0
NovpnFile	0.0201	0.9477	0	0	0	0	0.0008	0.0315	0	0	0	0	0	0
NovpnP2P	0.0115	0.0162	0.9415	0	0	0	0	0	0.0293	0.0016	0	0	0	0
NovpnStreaming	0	0	0	0.9485	0	0	0.0199	0.0125	0	0.0190	0	0	0	0
NovpnVoIP	0	0	0	0	1.0000	0	0	0	0	0	0	0	0	0
NovpnEmail	0.0046	0.0122	0	0.0074	0.0043	0.9391	0	0	0.0191	0.0134	0	0	0	0
VpnChat	0.0306	0	0	0	0	0.0085	0.9523	0.0085	0	0	0	0	0	0
VpnFile	0	0.0305	0.0176	0	0	0	0	0.9394	0	0.0125	0	0	0	0
VpnP2P	0	0	0	0.0095	0	0.0268	0	0	0.9637	0	0	0	0	0
VpnStreaming	0	0	0	0	0.0156	0	0.0266	0	0.0133	0.9446	0	0	0	0
VpnVoIP	0	0	0	0	0	0.0073	0	0	0	0	0.9846	0.0078	0	0
VpnE-mail	0	0.0029	0	0	0	0.0405	0	0	0	0	0	0	0.9565	0
NovpnChat														
NovpnFile														
NovpnP2P														
NovpnStreaming														
NovpnVoIP														
NovpnEmail														
VpnChat														
VpnFile														
VpnP2P														
VpnStreaming														
VpnVoIP														
VpnE-mail														

图9 识别结果的混淆矩阵

Figure 9 Confusion matrix of recognition results

图10展示了基于深度胶囊网络对在小数据集情况下加密流量识别结果的混淆矩阵,相较于大数据集,虽然识别的准确率有小幅下降,但在小数据集的条件下,本文模型能够取得较好的识别效果。

3.4.3 不同模型的准确率和效率对比与分析

不同模型准确率、精确度和召回率对比如图11所示,从中可以看出,较于传统的加密流量识别模型,DRCN模型识别的准确率、精确度和召回

率均是最高，分别为 0.964、0.969、0.963，比传统模型中效果最好的 BLSTM 模型分别提高了 0.055 4、0.066 7、0.058 4。

NovpnChat	0.947 8	0.011 0	0	0	0	0	0.041 2	0	0	0	0	0	0	0
NovpnFile	0.023 0	0.937 6	0	0	0	0	0.000 9	0.037 5	0	0	0	0	0	0
NovpnP2P	0.023 5	0.091 0	0.931 1	0	0	0	0	0	0.034 5	0.001 9	0	0	0	0
NovpnStreaming	0	0	0	0.944 6	0	0	0.021 4	0.013 5	0	0.020 5	0	0	0	0
NovpnVoIP	0	0	0	0	1.000 0	0	0	0	0	0	0	0	0	0
NovpnEmail	0.005 0	0.013 5	0	0.008 2	0.004 7	0.932 7	0	0	0.021 1	0.014 8	0	0	0	0
VpnChat	0.032 3	0	0	0	0	0.009 4	0.943 7	0.009 0	0	0	0	0	0	0
VpnFile	0	0.033 2	0.019 1	0	0	0	0	0.934 1	0	0.013 5	0	0	0	0
VpnP2P	0	0	0	0.010 6	0	0.029 8	0	0	0.959 7	0	0	0	0	0
VpnStreaming	0	0	0	0	0.016 0	0	0.027 3	0	0.013 7	0.943 0	0	0	0	0
VpnVoIP	0	0	0	0	0	0.007 9	0	0	0	0	0.983 8	0.008 4	0	0
VpnE-mail	0	0.005 5	0	0	0	0.076 2	0	0	0	0	0	0	0.918 5	0
NovpnChat														
NovpnFile														
NovpnP2P														
NovpnStreaming														
NovpnVoIP														
NovpnEmail														
VpnChat														
VpnFile														
VpnP2P														
VpnStreaming														
VpnVoIP														
VpnE-mail														

图 10 小数据集情况下加密流量识别结果的混淆矩阵
Figure 10 Confusion matrix of encrypted traffic identification results for small datasets

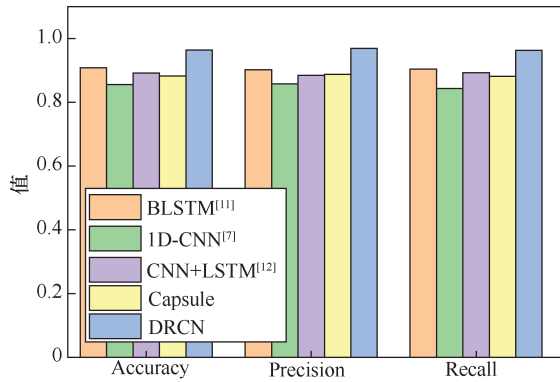


图 11 不同模型准确率、精确度和召回率对比
Figure 11 Comparison of accuracy, precision and recall of different models

不同加密流量识别模型的训练时间和检测时间对比如表 1 所示。由表 1 可知，网络结构最简单的 1D-CNN 模型的训练时间和检测时间是所有模型中最少的，但是由于其结构简单，识别准确率是最低的；DRCN 模型相较于基准的胶囊网络模型效率提高了 44%，在保持最高准确率的同时，训练和检测的时间较少，模型运行效率较高。DRCN 模型的效率较高是因为使用了 3D 卷积路由算法代替了基准胶囊网络全连接的路由算法，同时采用了残差连接的网络结构，能够有效提高模型的运行效率。

表 1 不同加密流量识别模型的训练时间和检测时间对比
Table 1 Comparison of training time and detection time of different encrypted traffic identification models

模型类别	训练时间/s	检测时间/s
BLSTM ^[11]	1 500	1.64
1D-CNN ^[7]	978	0.75
CNN+LSTM ^[12]	5 228	3.69
Capsule	2 865	1.83
DRCN	1 268	1.2

4 结束语

本文提出了一种基于深度胶囊网络的端到端加密流量识别方法。利用深度学习的优点避免了机器学习需要进行特征设计、特征提取等步骤，提高了识别的效率。在简单胶囊网络的基础上，使用 3D 卷积算法优化动态路由过程，实现了深度胶囊网络。同时引入通道注意力机制，加强了模型对重点特征的学习能力，提高了模型的识别效果。实验结果表明，本文模型比 1D-CNN 模型、BLSTM 模型和 Capsule 模型有更高的识别准确率，具有较好的应用前景。在未来的工作中有 3 个方面的问题需要解决：①采集更多的流量数据，平衡数据集；②模型对于某些流量的识别效果较差，如何提高这些流量的识别效果需要进一步研究；③考虑改进数据预处理方法，获得更好的模型输入数据。

参考文献:

- [1] AHMED M, MAHMOOD A N, HU J. A survey of network anomaly detection techniques[J]. Journal of Network & Computer Applications, 2016, 60:19-31.
- [2] DAINOTTI A, PESCAPE A, CLAFFY K C. Issues and future directions in traffic classification[J]. IEEE Network, 2012, 26(1): 35-40.
- [3] LUFT S J, CHIANG P. Network element architecture for deep packet inspection: US, US7719966 B2[P]. 2010.
- [4] ESTE A, GRINGOLI F, SALGARELLI L. Support vector machines for TCP traffic classification[J]. Computer Networks, 2009, 53(14):2476-2490.
- [5] KHONDOKE R, RAHAMATULLAH M. Interactive traffic class prioritization and radio link condition diversification techniques for efficient utilization of MMS and WAP services[J]. Chemical Communications, 2012, 48(98): 11960-11962.
- [6] QIU C, XU H, BAO Y. Modified-DBSCAN clustering for identifying traffic accident prone locations[C]//International Conference on

- Intelligent Data Engineering and Automated Learning. 2016: 99-105.
- [7] WEI W, MING Z, WANG J, et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks[C]//2017 IEEE International Conference on Intelligence and Security Informatics (ISI). 2017.
- [8] SHAPIRA T, SHAVITT Y. FlowPic: encrypted internet traffic classification is as easy as image recognition[C]// IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops. 2019.
- [9] REN X, GU H, WEI W. Tree-RNN: tree structural recurrent neural network for network traffic classification[J]. Expert Systems with Applications, 2021, 167(1):114363.
- [10] XIAO X, XIAO W, LI R, et al. EBSNN: extended byte segment neural network for network traffic classification[J]. IEEE Transactions on Dependable and Secure Computing, 2021 (1).
- [11] YAO H, LIU C, ZHANG P, et al. Identification of encrypted traffic through attention mechanism based long short term memory[J]. IEEE Trans Big Data, 2022, 8(1): 241-252.
- [12] ZOU Z, GE J, ZHENG H, et al. Encrypted traffic classification with a convolutional long short-term memory neural network[C]//2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City, IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). 2018: 329-334.
- [13] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 3859-3869.
- [14] RAJASEGARAN J, JAYASUNDARA V, JAYASEKARA S, et al. Deepcaps: going deeper with capsule networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10725-10733.
- [15] VENKATARAMAN S R, BALASUBRAMANIAN S, SARMA R R. Building deep equivariant capsule networks[C]//International Conference on Learning Representations. 2019.
- [16] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Training very deep networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2. 2015: 2377-2385.
- [17] 骆子铭, 许书彬, 刘晓东. 基于机器学习的 TLS 恶意加密流量检测方案[J]. 网络与信息安全学报, 2020, 6(1): 77-83.
- LUO Z M, XU S B, LIU X D. Scheme for identifying malware traffic with TLS data based on machine learning[J]. Chinese Journal of Network and Information Security, 2020, 6(1): 77-83.
- [18] 翟明芳, 张兴明, 赵博. 基于深度学习的加密恶意流量检测研究[J]. 网络与信息安全学报, 2020, 6(3): 66-77.
- ZHAI M F, ZHANG X M, ZHAO B. Survey of encrypted malicious traffic detection based on deep learning[J]. Chinese Journal of Network and Information Security, 2020, 6(3): 66-77.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.
- [20] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3d convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 4489-4497.
- [21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [22] DRAPER-GIL G, LASHKARI A H, MAMUN M S I, et al. Characterization of encrypted and VPN traffic using time-related[C]//Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP). 2016: 407-414.

[作者简介]



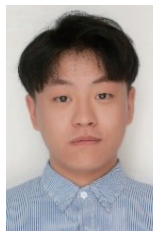
史国振（1974-），男，河南济源人，北京电子科技学院教授、博士生导师，主要研究方向为网络与系统安全、嵌入式系统。



李昆阳（1998-），男，江苏徐州人，北京电子科技学院硕士生，主要研究方向为信息安全。



刘瑶（1998-），女，陕西宝鸡人，西安电子科技大学硕士生，主要研究方向为信息安全。



杨永健（1997-），男，山东潍坊人，北京电子科技学院硕士生，主要研究方向为信息安全。