
WIRELESS NETWORK SECURITY DEEP LEARNING
WIRELESS TRAFFIC FINGERPRINT
PAPER PRODUCED BY GUET ZHANG HAN

声明

基于 CNN 的简单 APP 无线流量指纹识别（以下简称本项目）及其实验手册（以下简称本文），由桂林电子科技大学计算机与信息安全学院信息安全专业 2022 级张汉实现和编写。

有任何问题或学习交流，欢迎联系：hanz78843@gmail.com

注意！本文非严格的学术论文，而是基于交流学习而编写的。本着促进开放，交流学习的目的，项目源代码已上传：

https://github.com/27e7dyy38eu/Wireless-traffic-fingerprint-recognition-of-mobile-APP-based-on-simple-CNN

本项目及本文均遵循以下 MIT 开源协议

2025 年 5 月 16 日

摘要

随着移动互联网的高速发展，手机成为了生活中最重要工具，因此针对手机终端的隐私窃取行为也层出不穷。然而少有团队对手机无线加密流量进行分析。造成这一现象的原因有：1）不同设备、不同应用的无线流量混杂，难以区分；2）无线流量多为加密流量，难以进行分析；3）无线流量数量庞大，处理复杂度高；4）现有分析技术过于复杂，且多数不开源或仅声明开源，难以复现。因此，本项目在前人的基础上，提出了一种简单的、可复现的攻击方法：基于 CNN 的 APP 无线流量指纹识别。

本文第一章首先对项目进行一个简短的介绍，明确攻击的目的和对攻击进行界定；第二章则列出了读者需要具备的先验知识；第三章介绍了训练卷积神经网络模型使用的数据集；第四章详细的描述了攻击的步骤；第五章则呈现出实验的结果和数据；第六章对第五章的实验结果和数据进行讨论分析；第七章对整个实验项目进行总结，提出结论；第八章讲述了在本项目中的收获和感悟。

目前项目已实现 4 种手机 APP 的无线流量指纹识别。通过抓取 APP 启时的特征流量，在小批量数据集上的识别正确率超 97%。

关键词：隐私窃取、无线流量、卷积神经网络

目录

1 简介

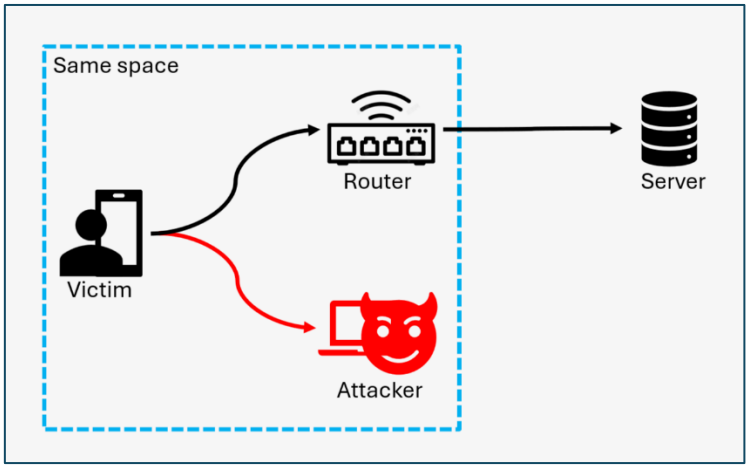

图 1-1 无线攻击场景

目前绝大多数流量分析攻击都是基于有线的攻击，即在报文传输链路上，如网线、路由器和被攻击机器等，直接进行抓包分析。然而在真实的攻击场景中，攻击者往往难以直接接触到被攻击设备，就不能进行抓包分析，因此有线攻击存在局限性。

而无线攻击就不需要直接接触受害设备，仅要求攻击者与被攻击者在同一信号可达空间内即可进行攻击[3]。同时这样的攻击场景普遍的存在与日常生活中，可以说无线攻击的适用场景无处不在。

因此，研究并实现一个简单的无线攻击是十分必要的，不仅证明攻击的可行性，更为未来的工作打下了坚实基础。

1.1 攻击目的

流量分析攻击的攻击目的都是获取流量中的隐私信息。本项目的攻击目的就是从加密的 802.11 帧中分析出被攻击设备正在或已经使用的 APP。

1.2 攻击行为划分

流量分析攻击不主动接触被攻击者，仅进行报文窃听、分析等工作，因此本项目属于被动攻击。

2 理论

2.1 IEEE 802.11

IEEE 802.11 是现今无线局域网通用的标准，是由电气和电子工程师协会（IEEE）所定义的无线网络通信的标准。虽然常将 Wi-Fi 与 802.11 混用，但二者并不相同。

设备之间进行无线传输前会选定一个固定信道，只要一次连接没有中断或结束，信道都不会改变。802.11 帧就是在链路层和物理层进行无线传输使用的报文，也就是本项目的分析对象。

2.1.1 802.11 帧分类

802.11 帧有三种：

1）管理帧：维护接入点和无线客户端之间的通信，有多种子类型；

2）控制帧：负责接入点和无线客户端之间的数据交换，有三种子类型；

3）数据帧：携带传输的数据的帧，即封装了上层报文，如 TCP、UDP 等，是本项目的主要研究对象。

由于管理帧和控制帧仅负责接入点和无线客户端之间的通信，他们本身不携带任何上层信息。只有数据帧承载上层报文，才与上层信息相关。但是数据帧的字型之一：NULL 型，也不携带任何信息。所以本项目所使用的报文都是除 NULL 型外的 802.11 数据帧。

2.1.2 802.11 数据帧格式

Frame Control	Duration ID	Address1 receiver	Address2 sender	Address3 filtering	Seq-ctl	Address4 Optional	Frame Body	FCS
2Byte	2Byte	6Byte	6Byte	6Byte	2Byte	6Byte	0-2312Byte	4Byte

图 2-1 802.11 数据帧格式

上层报文，都被封装在帧体 Frame Body 中。但是 802.11 帧在传输过程中帧体会被加密，因此无法直接从帧体得到任何信息。2.2 节介绍了一种特征表示，可以很好的描述某段流量的特征。第六章也讨论了这种特征表示在无线流量表征上也有很好的效果。

2.2 流量汇聚矩阵 TAM

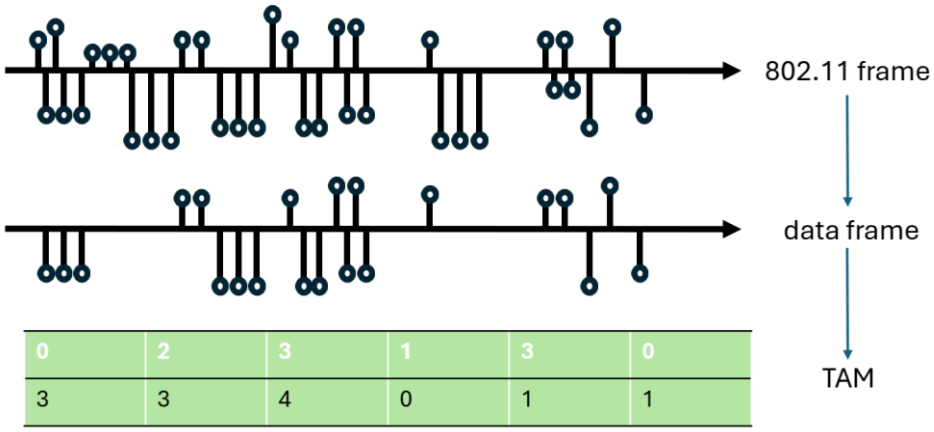

图 2-2 802.11 帧流量处理流程

北京理工大学的团队结合数据包的方向和时序特征，提出一种能高鲁棒表征网站流量指纹的流量表示方法——流量汇聚矩阵（TAM）[4]。

图 2-2 展示了如何将 802.11 帧流量处理成 TAM 的完整流程：

1） 过滤管理帧、控制帧、NULL 型数据帧；

2） 将一段时间内的帧分割为若干个小时间段，统计每个小时间段的上行帧和下行帧。

2.3 卷积神经网络 CNN

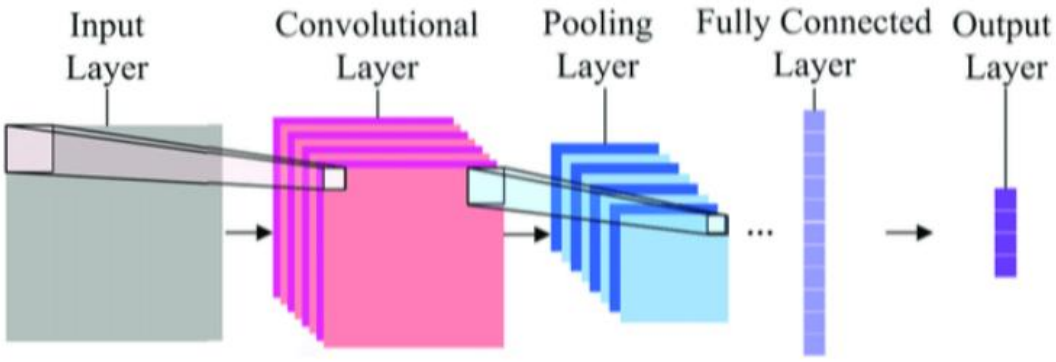

图 2-3 卷积神经网络结构

卷积神经网络（CNN），是一种在计算机视觉领域取得了巨大成功的深度学习模型。CNN 能很好的识别图片类别，对图像进行精确分类。

卷积神经网络（CNN）通过对图像进行卷积（Convolutional）、池化（Pooling）、全连接神经网络（Fully Connected）等 3 个层次的正向处理和反向传播，使模型学习到不同图像的特征模式，从而进行图像分类。

值得注意的是图像在计算机中以矩阵的形式存在，而 2.2 节中介绍的 TAM 正是$2 \times n$的矩阵，因此 CNN 也可以用来分类 TAM。

3 数据集

本项目的数据集由作者采集制作，经过数据帧过滤、优质帧筛选，包含 10 个手机 APP 各 50 条，共 500 条启动流量。包含 802.11 数据帧及对应 TAM。

采集的 APP 无线流量有：微信、淘宝、哔哩哔哩、支付宝、QQ 浏览器、美团、微博、京东、QQ、抖音。

自制数据集的方法与第四章 4.1 节的流量采集方法一致。

4 实验方法

本项目的实验环境如下表所示：

设备名	类型	版本号
Kali	虚拟机、攻击机	~
H3C Magic	路由器	~
Huawei mate70	手机、被攻击机	~
WiFi6 kali 专用 usb 网卡	外接带 monitor 模式网卡	~

4.1 数据采集

传统有线攻击使用五元组<Src IP,Dst IP,Sr port,Dst port,Protocol>来唯一标识一个报文。但是无线流量由于加密和层级不同，不能得到这些信息，且报文可能分片、混杂，因此需要构造新的元组来描述无线流量，以便进行抓包。

本文提出了六元组< Channel,Src MAC,Dst MAC,Protocol,start time,end time>来描述 APP 在某一时间段内的产生的无线流量。

由于设备原因，本实验抓取的流量均是在 2.5G 频段下的。

使用 kali 中的 wireshark 进行无线流量抓取，**在流量不进行变化时进行抓包，尽量保证抓取的流量的第一个 802.11 帧由 APP 产生或与 APP 产生的第一帧的时间间隔较小。**

流量采集的操作命令已在附录给出。

4.2 数据处理

图 2-2 展示了将采集到的流量转换为 TAM 的全流程。

由于抓取流量时只抓取与被攻击设备 MAC 有关的帧，因此通过源地址是否是被攻击设备的 MAC 就可以区分上下行数据包。

数据处理参数已在附录给出。

4.3 CNN 模型训练

4.3.1 模型结构

模型简单结构如下所示：

1） Conv1

2）Conv2

3）Pool

4）Dropout

5）Linear1

6）Linear2

4.3.2 模型参数

模型参数已在附录中给出。

5 结果分析

5.1 相似性分析

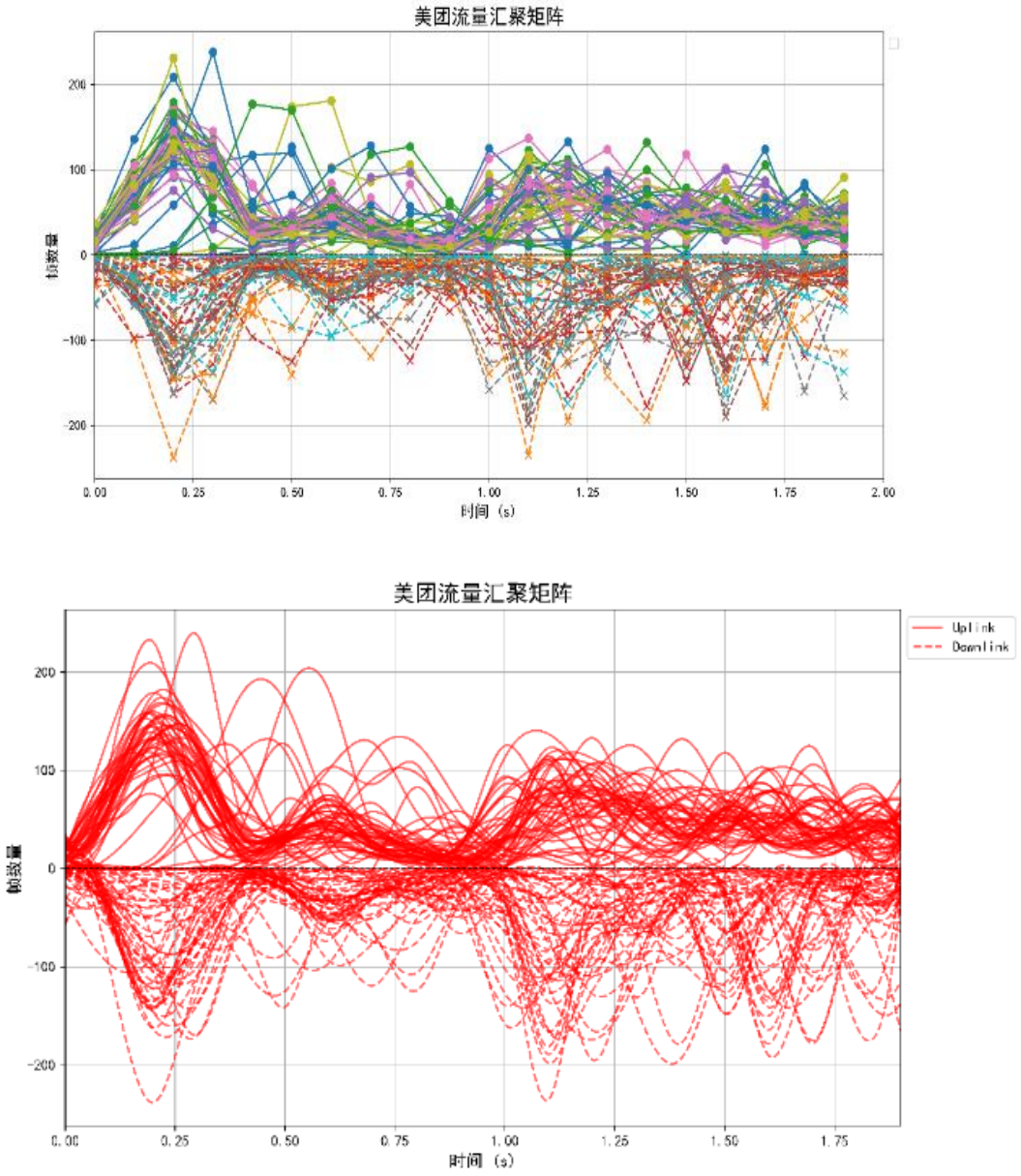

图 5-1 美团的流量汇聚矩阵及其平滑化

图 5-1 展示了美团启动时前 2s 内的 TAM，共 50 条流量，显然这 50 条流量的 TAM 具有一定的相似性。

5.2 可分性分析

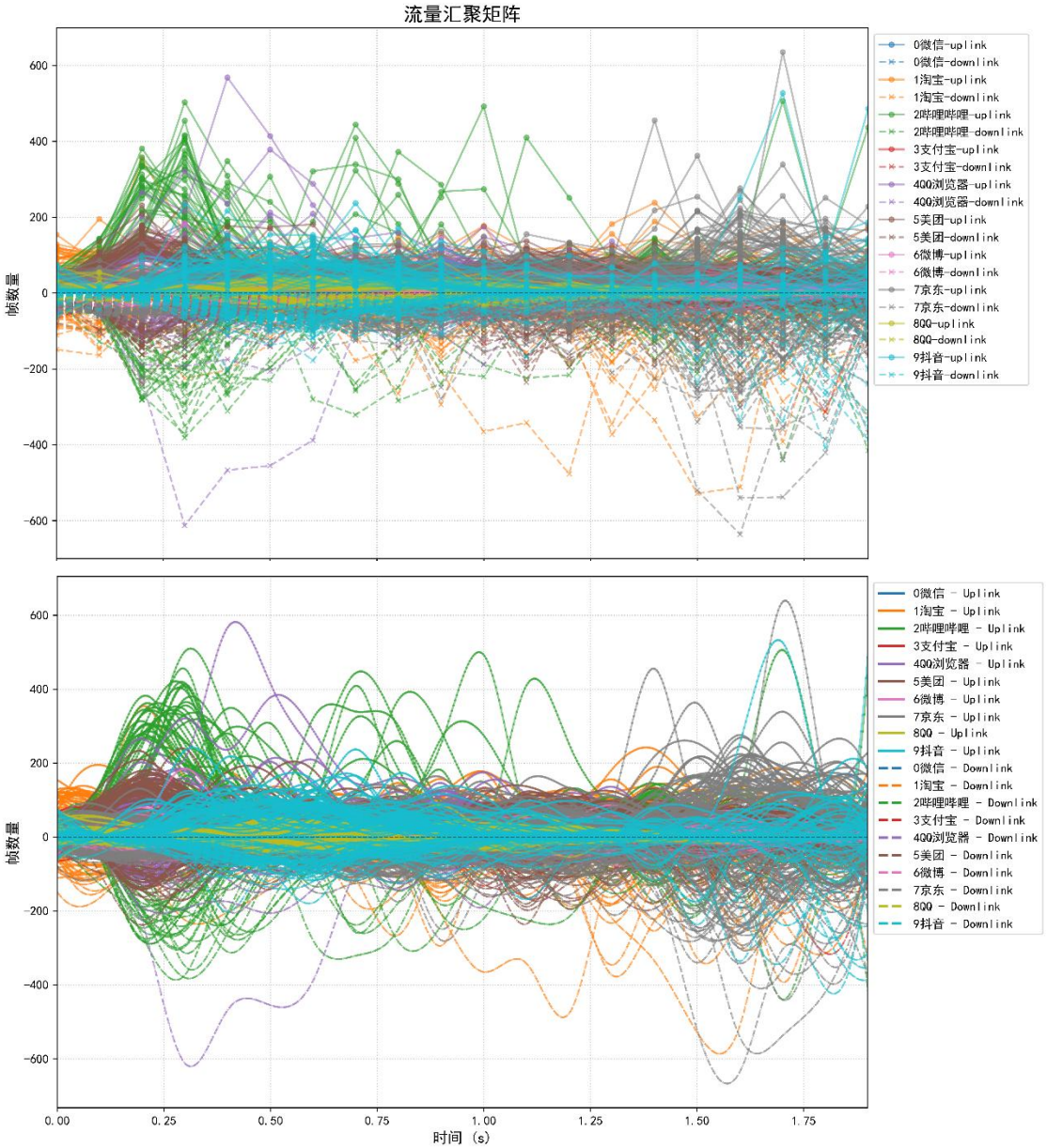

图 5-2 10 个 APP 的流量汇聚矩阵及其平滑化

图 5-2 展示了 10 个 APP 无线流量 TAM 的对比，显然这 10 个 APP 的 TAM 的最大值、活跃时间段和总数极不相似，可以初步确定这 10 个 APP 无线流量在 TAM 上具有一定的可分性。

5.3 分布规律

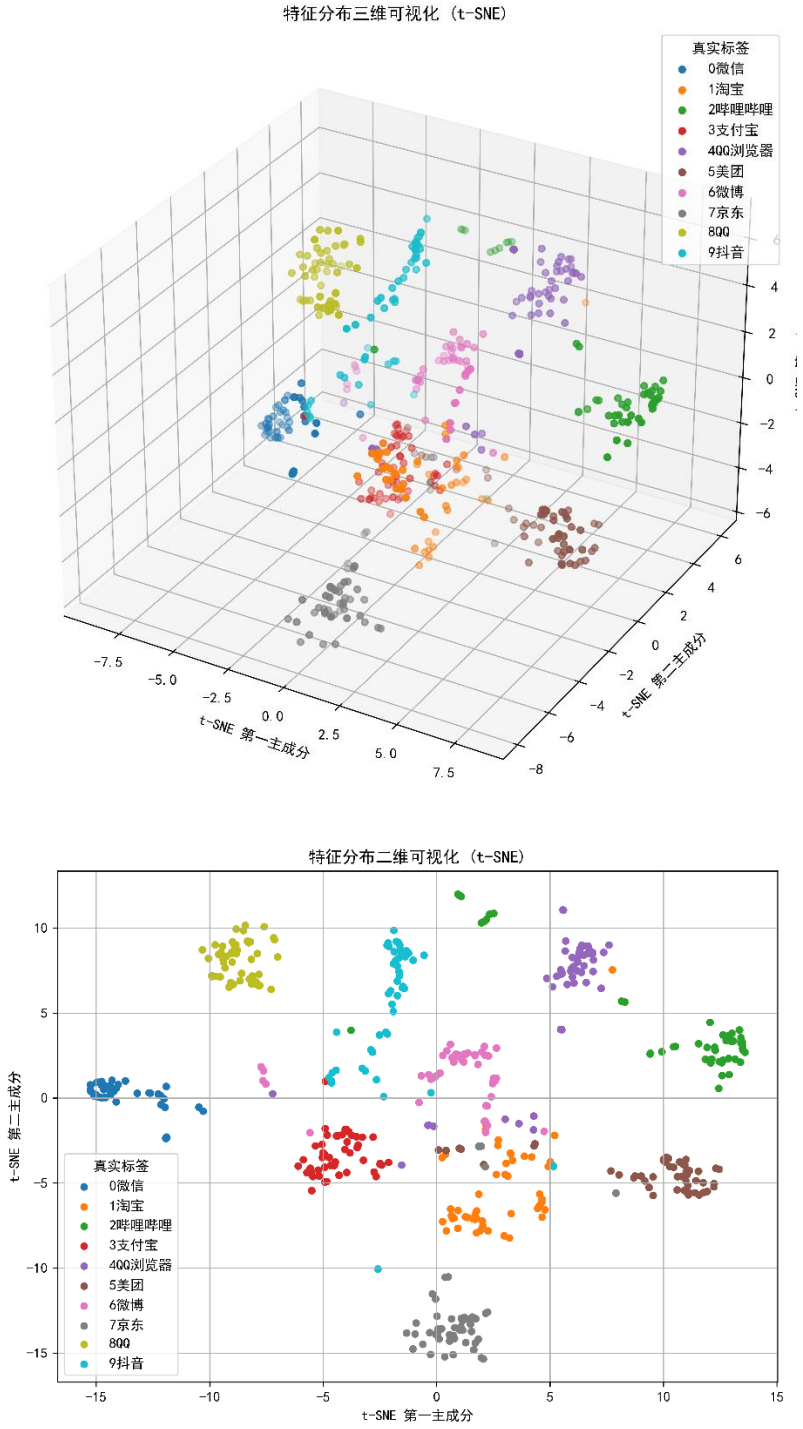

图 5-3 10 个 APP 分类后三维及二维空间分布

将训练集作为输入，重新输入至训练好的 CNN 模型中，使用 t-SEN 方法将其映射至低维空间，可以观察到 CNN 在该数据集上学习到的分布规律。图 5-3 描述了训练集分类后在三维及二维空间中的分布，表明 CNN 模型在这一数据集上确实有一定的分类效果。

5.4 模型评价指标

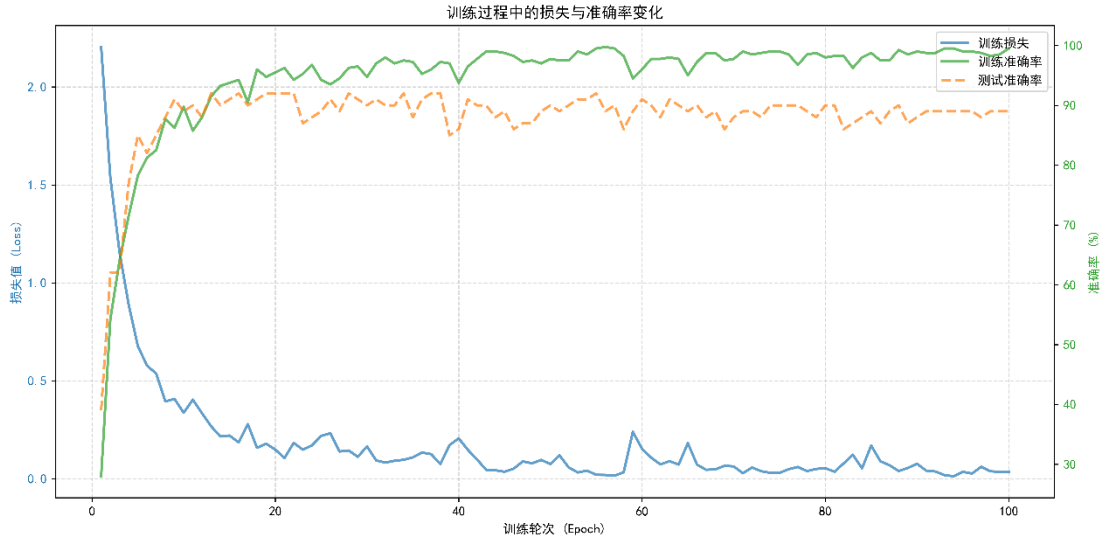

图 5-5 模型训练正确率、测试正确率和损失

图 5-5 描述了 CNN 模型在 100 论训练内的训练正确率、测试正确率均收敛于较高水平，同时损失也收敛于低水平。

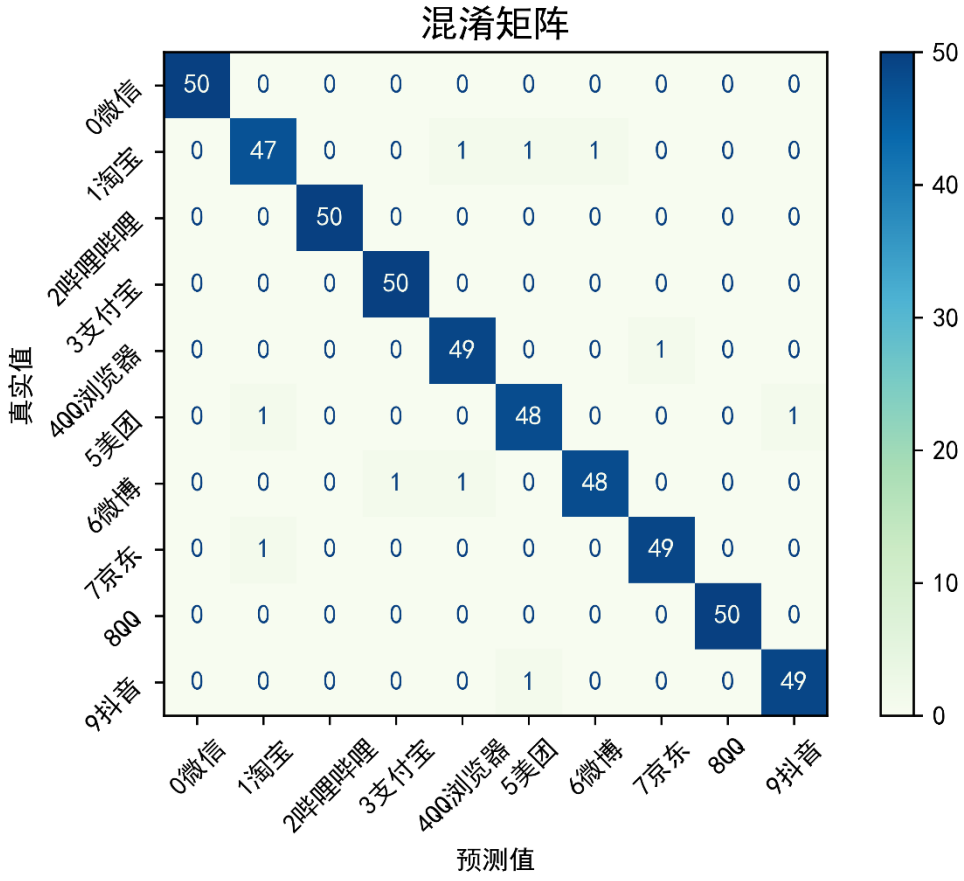

图 5-6 CNN 模型混淆矩阵

图 5-6 的混淆矩阵描述了该 CNN 模型训练的最终效果。能将绝大部分的无线流量正确的对应至其标签。

6 结论

通过实验及结果分析，本文得出以下结论：

1) 这 10 个 APP 的无线流量在 TAM 上确实是可分的；

2) 这 10 个 APP 的无线流量可以使用简单 CNN 模型进行分类；

3) 无线流量是可以分类的。

未来工作：

1） **更多更棒！**由于技术原因，目前作者制作的数据集包含很多噪声，同时数量对于一般的 CNN 数据集来说也略少。因此需要构建更大批量优质的数据集；

2） **更大更好！**本项目使用的模型是深度学习中最最简单的模型，未来可以尝试参数量更大和机制更优秀的模型；

3） **更准更快！**项目目前并不能达到作者预想的实时检测效果，即实用工具，因此未来将会专注于实时检测系统的工作。

7 后话

这只是一个很小很小的微不足道的项目，但后话不是这个项目的终点。或许它不如什么毕设、期刊，但是这是我收获最大的项目，是很值得我写这一段文字的。

其实我做这个小项目只有一个目的——我就想做点不一样的东西。

我不想做人家现成的东西，我就想做别人没有做过的，别人做不成的，做点我自己的东西。但是这条路实在是太艰难了，从开始的四人团队，慢慢的仅剩我一人。每每看着毫无规律的数据时，我都不住的怀疑，这个东西真的能做成吗？每每查询相关文献却没有多少时，我都不住的怀疑，这个东西真的能做成吗？每每复现不出他人的实现结果时，我都不住的怀疑，这个东西真的能做成吗？每每深夜失眠，我不止一次想要放弃，多少次想着，就这样吧，算了吧。但是心里总憋着一股劲。这股劲的源头，是一年前大二的一堂“水课”，是一门《网络安全法》，是一个老师的点拨。正是因为这股劲，我在低落后总能打起精神，继续上路。

我面临的最大的困难，不是论文看不懂，不是代码不开源，而是我自己。总是害怕失败而不去尝试，万一做不成呢？那我的时间不是浪费了吗？那还不如去打打游戏。我也不清楚我是怎么坚持下来的，只是默默埋首向前。

所幸，取得了一些微不足道的进展，于是有了这篇文章。成功的那一刻，才明白，最重要的不是结果，而是经历过。只有经历过才能明白，夤夜思考却不得法门的折磨痛苦，灵光乍现的欣喜若狂。一次小小的成功微不足道，但却给了我极大的信心。研究的路上总是需要一些正反馈来作慰藉，才能走得更远。

这个项目能成功，首先要感谢我的导师曹老师，在立项之初就给予了我很多帮助和支持，特别是遇到困难时在办公室一起讨论，我都能从曹老师身上学到很多。研究认真严谨，做事负责，乐观豁达。还要感谢姚老师在数据采集上的点拨和黄老师为我解答无线安全领域的一些困惑。

感谢各位老师在这个项目上对我的帮助和指点，再多的汗水也比不上刹那的灵感！以及各位给我提供情绪价值的朋友们！没有你们的帮助就没有这个小项目。我衷心感谢你们！

最后的最后，赞美知识！赞美坚持！赞美那些坚持开源的研究者们！

参考文献

[1] Carolin Svensson.Anomaly Detection in Encrypted WLAN Traffic[D]. Linköping:Linköping University,2020.https://www.researchgate.net/publication/354742459_Anomaly_Detection_in_Encrypted_Internet_Traffic_Using_Hybrid_Deep_Learning.

[2] Jianfeng Li.Packet-Level Open-World App Fingerprinting on Wireless Traffic[C]. NDSS,2022.https://dev.ndss-symposium.org/wp-content/uploads/2022-210-paper.pdf.

[3] Jian Qu.An Input-Agnostic Hierarchical Deep Learning Framework for Traffic Fingerprinting[C].USENIX,2023.https://www.usenix.org/conference/usenixsecurity23/presentation/qu.

[4] Meng Shen.Subverting Website Fingerprinting Defenses with Robust Traffic Representation[C].USENIX,2023.https://www.usenix.org/conference/usenixsecurity23/presentation/shen-meng.

附录

1 流量采集操作命令表（kali）

sudo airmon-ng start wlan0	改变网卡为 monitor 模式
sudo airodump-ng wlan0mon	扫描附近热点信道
sudo iw dev wlan0mon set channel 6	设置网卡监听信道
Wireshark 过滤语句： (wlan.addr==路由器 MAC && wlan.addr==手机 MAC &&wlan.fc.type==2)&&!(wlan.fc.subtype==4)&&!(wlan.fc.subtype == 12)	

2 特征提取参数

参数名	**值**
T	2s
S	0.1s

3 模型训练参数

参数名	**值**
input_channels	**2**
max_length	50
batch_size	4
epochs	100
learning_rate	0.001
optimizer	Adam