基于深度学习的蓝牙射频指纹识别系统

陈 拓 杨 洁 翟宇辰 安晨珲 李宗岩
（南京工程学院 南京 210000）

摘 要 蓝牙射频指纹具有难以伪造的优点，基于射频指纹的身份识别能有效提高网络的安全性。文中设计了一种基于深度学习网络的蓝牙射频指纹识别系统。首先，利用 Hackrf One 软件无线电平台和 GNU Radio 软件在蓝牙信号广播阶段采集多种蓝牙信标信号。其次，对蓝牙信号进行预处理，将预处理后的数据分为训练集与验证集。然后，使用 MATLAB 深度学习工具箱来设计长短期记忆网，利用训练数据集对各个网络进行训练，得到蓝牙射频指纹识别网络。最后，利用验证集对上述网络进行测试和分析。当迭代次数为 300 时，网络对 3 种蓝牙信标的射频指纹识别的准确率均达到 80％以上。

关键词：蓝牙；软件无线电；射频指纹识别；深度学习；LSTM
中图法分类号 TN915.08

Bluetooth Radio Frequency Fingerprint Identification System Based on Deep Learning

CHEN Tuo，YANG Jie，ZHAI Yuchen，AN Chenhui and LI Zongyan
（Nanjing Institute of Technology，Nanjing 210000，China）

Abstract Bluetooth RFID fingerprint has the advantage that it is difficult to forge，and the identification based on RF fingerprint can effectively improve the security of the network. A Bluetooth RF fingerprint identification system based on deep learning network is designed in this paper. First，use the Hackrf One software radio platform and GNU Radio software to collect a variety of Bluetooth beacon signals in the Bluetooth signal broadcasting stage. Secondly，the Bluetooth signal is preprocessed，and the preprocessed data is divided into training dataset and validation set. Then，the MATLAB deep learning toolbox is used to design the long short-term memory network，and the training data set is used to train each network to obtain the Bluetooth radio frequency fingerprint recognition network. Finally，the above network is tested and analyzed by using the validation set. When the number of iterations is 300，the accuracy rate of the radio frequency fingerprint recognition of the three Bluetooth beacons by the network reaches more than 80％.

Keywords Bluetooth，Software defined radio，RFF（Radio Frequency Fingerprinting），Deep learning，LSTM

0 引言

近年来，由机器学习演变而来的深度网络与信号射频指纹识别结合的技术逐渐成熟[1]。蓝牙指纹识别技术也从需要通过复杂的人工辨别处理，转变为仅需通过深度学习提取蓝牙指纹特征，完成分类识别[2]。蓝牙 5.0 具有低功耗、高传输速率、长传输距离等优点，得到了广泛的应用。

传统的以密钥认证技术为核心的安全技术已不能识别蓝牙信标身份，基于简单数字标识符（如硬件地址、信标序列号）的标识解决方案也无法有效识别信标身份。射频指纹是蓝牙信号发射机的内部元器件在生产和工作过程中存在的不可避免的误差对产生的射频信号造成的独特影响[3]，这种干扰和噪声与设备本身紧密相关，具有独一无二的特征。目前，射频指纹识别技术仍存在一定的不足[4]。一方面，现有研究主要集中于 WLAN 射频指纹，对蓝牙射频指纹识别的研究较少；另一方面，目前基于深度学习的射频指纹识别方法的参数较多，计算量较大，在各类应用场景中的可行性较低[5]。

收稿时间：2023-08-03

作者简介：陈拓（2001—），本科生，研究方向为无线通信。

1 系统总架构

本文研究了一种基于 LSTM 的蓝牙射频指纹识别系统，该系统将深度学习网络引用到蓝牙信号识别上，利用 Hackrf One 采集到的原始数据信号直接提取信号特征并完成身份识别[6]。该系统在保证蓝牙指纹识别准确率的基础上，减少了数学分析环节，提高了蓝牙指纹识别性能[7]。本文利用软件定义无线电 Hackrf One 采集数据，并建立数据集，具体方法为如下。

在 VM 虚拟机的环境下使用 GNU Radio 软件对 3 种

不同型号的蓝牙信标信号进行数据采集，并将数据以十六进制的形式保存下来。在预处理过程中，将十六进制的数据样本转化为十进制数据，组成训练集和验证集矩阵后，将其导入 MATLAB 中。接着将经过预处理后的信号训练集与验证集分别导入 MATLAB 中进行验证与测试，从而实现身份分类识别。

2 蓝牙信号的采集

2.1 GNU radio 信号接收系统设计

本文在基于 VM 虚拟机的环境下使用 GNU Radio。GNU Radio 具有高度模块化、基于流程图等特点，针对复杂的信号处理应用场景，其提供了丰富、全面的处理模块。Hackrf One 是一种软件定义的无线电，可以快速、准确地传输无线电信号。它也是一个开源平台，可用作 USB 的外围外设。本文在 GNU Radio 操作过程中使用了 QT GUI 图形用户界面，搭建的信号采集系统如图 1 所示。

本文使用 3 类不同型号的蓝牙信标，GNU Radio 的采样频率 Samp_rate 为 10MHz/s。在采集蓝牙信号数据时，考虑到低功耗 BLE 的广播信道为 37，38，39 号信道，3 个信道的中心频率分别为 2 402MHz，2 026MHz，2 480MHz。本文根据不同信道的中心频率更改了 Osmocom Source 蓝牙信号接收模块的中心频率，以排除频率混叠造成的干扰。

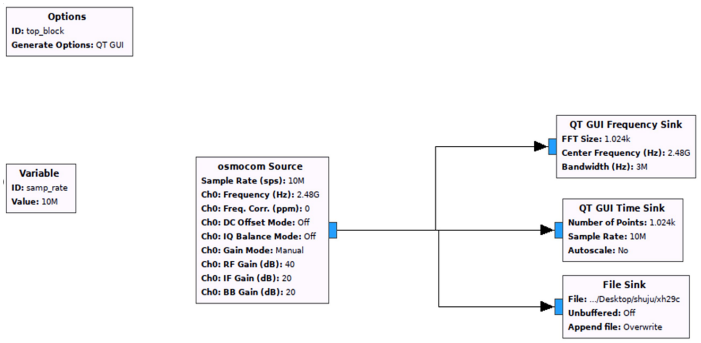

图 1 蓝牙信号采集系统

2.2 采集蓝牙信号

以 E-BCN05L 型号为例，在本系统的设计中，该信号的信标广播信道为 38 信道，中心频率为 2 426MHz。以 E-BCN05L 型号信标为例，采集的蓝牙信号的时域、频域波形如图 2 所示。时域波形图中的红色和蓝色线条分别代表信号的实部和虚部。

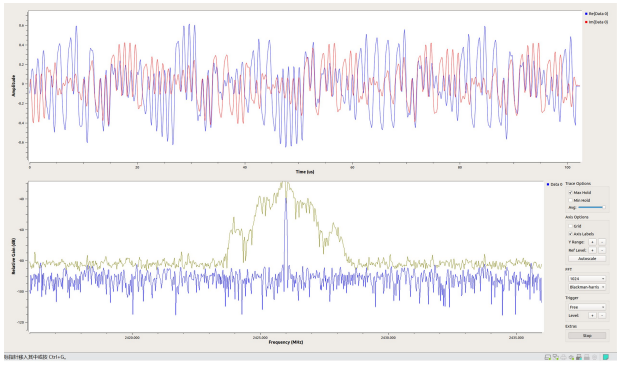

图 2 E-BCN05L 的时域、频域图（电子版为彩图）

之后，将捕捉到的波形数据在 Utubun 系统中使用命令行以十六进制的形式查看。

3 蓝牙信号预处理

由于各应用系统的数据缺乏统一的标准和定义，数据结构存在较大的差异。深度学习算法虽然具有很强的特征抓取能力和学习能力，但深度学习算法对数据的要求很高，因此需要对蓝牙射频信号进行有效的预处理，使其具有物理意义或统计特征，便于后续识别。具体的处理方法如下。

收集 3 类蓝牙信标数据（每类 3 组），共有 9 组数据，将这 9 组以十六进制方式保存的波形文件进行处理（先进行删 0），再转为十进制的保存形式。在每类信标的 3 组数据中各随机抽取 50 行数据，拼接为一个 150×16 的矩阵形式，并以 Excel 形式保存下来，作为这一类信标的训练样本。之后，在每类信标的 3 组数据中各随机抽取 30 行数据，将其拼接成一个 90×16 的矩阵，并以 Excel 的形式保存下来，作为这一类信标的测试样本。

4 深度学习网络的选型与设计

4.1 网络选型

本文使用 MATLAB 深度学习工具箱来设计长短期记忆网络，利用训练数据集对各个网络进行训练，得到蓝牙射频指纹识别网络[8]。LSTM 有很多优势，它适用于学习时间序列，仅需要较少的辅助处理环节，在时间学习方面具有显著的效果[9-10]。

4.2 LSTM 网络的搭建

4.2.1 生成训练集

由于 LSTM 网络的特殊性和对输入数据集的要求，LSTM 网络要求网络的输入数据类型必须是 Cell（细胞型数据），因此需要使用 Mat2cell 函数将矩阵类型的数据转为 Cell 型。同时，标签集也要相应地转化为 Cell 型。

4.2.2　LSTM 网络训练

序列输入层是 LSTM 网络的独特输入层，输入的数据是时序数据集，保证了网络在时序中对数据集进行分类。LSTM 层的作用是学习数据集中的时间序列和步长之间的依赖关系，并帮助 LSTM 深度学习网络在训练期间深化长序列上的梯度关系。

4.2.3　分类测试部分

具体代码如图 3 所示。由于 LSTM 网络的特殊性和对输入数据集的要求，其对于验证集和验证集标签的数据类型也有不同的要求，即要求必须是细胞型数据。

```
cs1=readmatrix('H1.xls','Sheet','Sheet1','Range','A151:P300');
cs2=readmatrix('H1.xls','Sheet','Sheet2','Range','A151:P300');%cs1\cs2\cs3分别为三类信标的测试样本
cs3=readmatrix('H1.xls','Sheet','Sheet3','Range','A151:P300');%并从EXCEL中导入数据
Xtest1=[cs1;cs2;cs3];                                    %三类样本拼接成90*16的矩阵，组成训练集
rowDist2 = [150 150 150];
Xtest = mat2cell(Xtest1,rowDist2);


ya=zeros(1,1);                                           %零矩阵处理作为标签
yb=ones(1,1);                                            %一矩阵处理作为标签
yc=ones(1,1)*2;                                          %二矩阵处理作为标签
ytest1=[ya;yb;yc];                                       %标签测试集
ytest=categorical(ytest1);                               %函数包要求标签类型是categorical

YPred = classify(net,Xtest);                             %网络测试
YPred1 =double(YPred);                                   %转化为可显示的标签
accuracy = sum(YPred == ytest)/numel(ytest)              %精确度计算
```

图 3　产生验证集及网络测试部分代码

5　实验与分析

5.1　网络迭代测试

3 类蓝牙信标都会收集 3 组数据，共有 9 组数据，参与训练的样本集共有 450 个信号样本，是一个 450×16 的矩阵。验证集有 90 个信号样本，是一个 90×16 的矩阵。LSTM 网络测试结果显示，在迭代次数为 300 次时，验证集中的合法信号帧为 87 个，准确度为 0.966 7。训练、验证结果如图 4 所示。

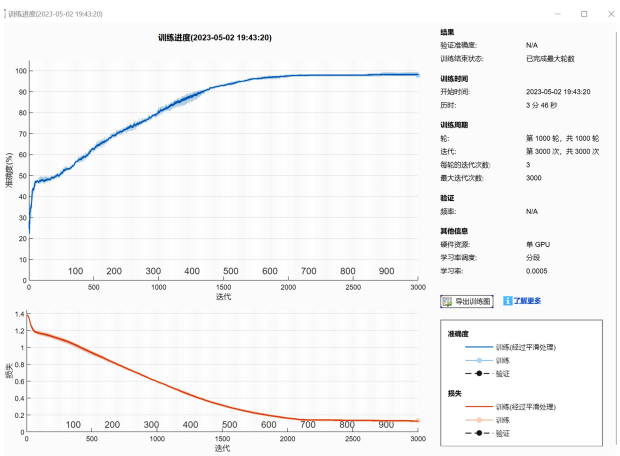

图 4　迭代次数为 300 时 LSTM 网络训练、验证结果

5.2　测试系统

(1)训练集分类结果对比

训练集经过 LSTM 网络深度学习之后，用训练集测试网络，对分类结果进行对比分析，如图 5 所示。

由图 5 可知，训练集分类结果对比准确率为 94.5%，如在第 1 个设备的分类对比中，系统会根据训练提取到的设备的特征，对数据进行预测分类，如第 28 个数据的分类值为 2，即训练后的系统预测的第 28 个数据是由第 2 个设备发出的，但事实上其却是由第 1 个设备发出的，存在一定的误差。但 5.5%的误差尚在可接受范围之内。

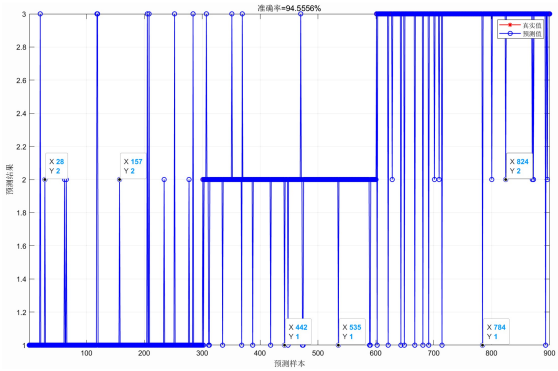

图 5　训练集分类结果对比

(2)测试集分类结果对比

用测试集测试网络，并对分类结果进行对比、分析，结果如图 6 所示。

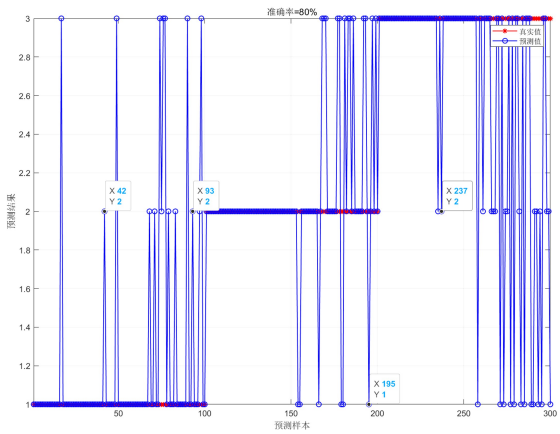

图 6　测试集分类结果对比

测试集分类结果的准确率为 80%，由于信号采集环境复杂等原因，测试集分类的准确率较训练集而言，存在一定的下降，但也可以达到 80%以上，预测准确率较为理想。改进算法、提高准确率，是下一步的研究重点。

6　结语

本文利用 GNU Radio 和 Hackrf One 在蓝牙信号广播阶段采集信号，并形成数据集，通过对数据集进行深度学习，在物理层方面有效应对来自冒充设备的威胁，提升了无线通信网络的安全性[11]。基于该思想，本文设计了基于 LSTM 的蓝牙射频指纹识别系统。实验结果表明，该系统可以实现对蓝牙射频指纹的有效识别。基于深度学习的射频指纹识别技术，可以更好地应对未来无线通信网络发展所带来的通信安全问题。同时，本文在深度学习网络的选取上还可以进一步提高，且缺少在更加复杂的情况下的验证情况。

参考文献

[1] 曾勇虎，陈翔，林云，等.射频指纹识别的研究现状及趋势[J].电波科学学报，2020，35(3)：305-315.

[2] 俞佳宝.射频指纹提取与识别技术研究[D].南京：东南大学，2020.

[3] 张辉，范雪林，李东风.一种改进的短时傅里叶算法实现[J].通信技术，2017，50(5)：1066-1069.

[4] 张振，贾济铖，康健，等.射频指纹识别技术方法综述[J].无线电通信技术，2021，47(3)：249-258.

[5] 贾济铖，齐琳.基于双谱的射频指纹提取方法[J].太赫兹科学与电子信息学报，2021，19(1)：107-111.

[6] 张驰，张峰，刘叶楠，等.基于融合聚类的蓝牙指纹室内定位算法优化[J].计算机仿真，2020，37(7)：314-318.

[7] 张谦，王吉，唐泽宇，等.基于深度学习的通信电台个体识别技术[J].电子信息对抗技术，2021，36(2)：36-40.

[8] D.W，NAGATA Y，SANO M，et al.A Scanner Error Discriminator Based on Short-time Fourier Transform in Pulse Train Interferometry[J].Optics Communications，2021，488(1)：126816.

[9] 张辉，范雪林，李东风.一种改进的短时傅里叶算法实现[J].通信技术，2017，50(5)：1066-1069.

[10] J.Y.LU.A New Indoor Location Algorithm Based on Radio Frequency Fingerprint Matching[J].IEEE Access，2020(8)：83290-83297.

[11] 刘念.基于射频指纹的无线网络安全增强机制[D].北京：北京交通大学，2020.

(上接第 214 页)

据集成、规范设计、数据溯源、数据转换、任务调度等功能，可支撑数据模型的全生命周期管理。

5.2.7 可视化展示

可视化展示采用图表和页面动画的形式，能更加直观地理解和分析数据。展示的数据可以通过数据库和文件上传两种方式进行增删查改，做到实时、动态的更新。

6 结语

以政府需求为导向，地市级能源大数据中心可充分发挥电力大数据优势，参考省级能源大数据中心模式，利用现有的数据初步打造应用场景，支持当地社会经济的发展。同时，多措并举推进水、电、煤、油、气、热等多种能源数据的汇聚融合，逐步分析重点行业和区域碳排放、碳减排、碳中和、碳交易等数据，打造一站式的“能源＋双碳”数据中心。下一步，地市级能源大数据中心将在相关部门的指导下，持续提升运营能力、创新业务模式，以服务“双碳”为目标，综合各方需求，充分发挥地缘优势，挖掘具有地方特色的能源大数据产品，推动能源大数据中心的进一步发展。

参考文献

[1] 都兰娜，寺林嚞.“双碳”视角下的电网企业能源数字经济发展路径[J].中国电力企业管理，2022(4)：72-73.

[2] 李飞.基于数据中台的智能工厂信息化平台设计[J].电脑知识与技术，2023，19(3)：57-59.

[3] 胡学强.基于大数据挖掘的电力客服中台数据智能整合方法[J].计算机与通信技术，2023，42(3)：117-121.

[4] 李飞.基于数据中台的智能工厂信息化平台设计[J].电脑知识与技术，2022，20(12)：142-145.