基于自适应滤波算法的有线网卡指纹提取方法

胡园园[1], 胡爱群[1], 李 晟[2], 刘佳琪[3] 李 冰[1]

[1] 东南大学网络空间安全学院 南京 中国 211189
[2] 东南大学信息科学与工程学院 南京 中国 211111
[3] 南京理工大学紫金学院计算机学院 南京 中国 210023

摘要 有线设备接入认证是保障有线以太网安全的重要组成部分, 其中 MAC 地址认证和设备数字证书认证是目前的主流身份认证方式, 然而前者存在 MAC 地址易被篡改和伪造, 后者存在系统复杂、使用不便等问题。基于设备指纹的物理层安全技术是解决这一问题的有效途径, 并已在无线网络中得到广泛应用, 但有线网络目前研究颇少。设备指纹的提取是物理层安全技术的一个重要环节, 有线网络已有研究主要从 10M 有线网卡信号中提取指纹。本文提出了一种基于最小均方误差自适应滤波算法(LMS 算法)从 100M 有线网卡信号中提取指纹的方法, 该方法提取的网卡指纹产生自网卡及所在设备本身的物理特性, 不可克隆, 无法被篡改, 而且指纹可直接通过分析网卡输出信号而得, 简单方便。本文设计了一套基于 LMS 算法的网卡指纹提取系统, 通过大量实验估算了合适的诸如收敛因子、滤波器阶数、数据长度等算法参数, 并对提取的指纹进行了有效性验证。经过实验验证, 使用本文方法提取的网卡指纹可有效识别出不同品牌和相同品牌不同类型的以太网网卡, 在使用线性判别和集成子空间判别分类算法时, 针对 50 块网卡的识别率可分别达到 97.3%、98.5 以上。

关键词 有线以太网网卡, 指纹提取, LMS 算法, 身份认证

中图法分类号 TP309.1 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2022.07.10

Fingerprint extraction of Ethernet card based on adaptive filtering algorithm

HU Yuanyuan[1], HU Aiqun[1], LI Sheng[2], LIU Jiaqi[3], LI Bing[1]

[1]School of Information Science and Engineering, Southeast University, Nanjing 211189, China
[2]School of Cyber Science and Engineering, Southeast University, Nanjing 211111, China
[3]School of Computing, Nanjing University of Science and Technology Zijin College, Nanjing 210023, China

Abstract The identity authentication of access wired devices is an important part of the security of wired Ethernet, among which MAC address authentication and digital certificate authentication of the device are the mainstream authentication methods at present. However, the MAC address in the former authentication is easy to be tampered and forged, while the latter has problems such as complex system and inconvenient to use. Physical layer security technology based on device fingerprint is an effective way to solve these problems, and has been widely used in wireless networks, but there is little research on wired networks. Device fingerprint extraction is an important part of physical layer security technology. Present studies of physical layer security technology in wired network mainly extract fingerprint from 10M wired Ethernet card signal. This paper proposes a method to extract fingerprint from 100M wired Ethernet card signal based on the least mean square error adaptive filtering algorithm (LMS algorithm). The fingerprint extracted by this method is generated from the physical characteristics of the Ethernet card and the device where the Ethernet card resides, and cannot be cloned or tampered with. Moreover, the fingerprint can be obtained directly by analyzing the output signal of the Ethernet card, which is simple and convenient. This paper designs a wired network card fingerprint extraction system based on LMS algorithm, estimates the appropriate algorithm parameters (such as convergence factor, filter order, data length and so on) through a lot of experiments, and verifies the validity of the extracted fingerprint. Experimental results show that the network card fingerprint extracted by this method can effectively identify wired Ethernet cards of different brands and different types of wired Ethernet cards of the same brand. When linear discrimination classification algorithm and integrated subspace discrimination classification algorithm are used, the recognition rates of 50 wired Ethernet card can reach 97.3% and 98.5 respectively.

Key words wired Ethernet card, fingerprint extraction, LMS algorithm, identity authentication

---

通讯作者: 胡爱群, 博士, 教授/博导, Email: aqhu@seu.edu.cn。

本课题得到江苏省重点研发计划“电力物联网边缘接入安全技术研究与应用”项目(No.BE2019109)资助。
收稿日期: 2021-08-14; 修改日期: 2022-01-20; 定稿日期: 2022-05-11

1 引言

对终端设备实施接入认证以杜绝非法接入是保障有线网络安全的重要措施。目前应用的接入认证方法主要有对终端设备的 MAC 地址进行认证, 以及基于数字证书的安全认证。但 IP 地址和 MAC 地址容易被篡改和伪造, 因此基于 MAC 地址的接入认证不足以保障有线网络的安全。而基于数字证书的接入认证虽然相对安全, 但使用过程复杂且数字证书本身也存在被窃取、盗用的风险, 这在大规模的网络应用中是不方便的。因此, 有线网络需要更安全和方便的安全机制。

近些年的研究表明, 通信用设备发射的信号具有物理指纹特征。该特征类似于人的指纹, 称为设备指纹或设备 DNA, 是设备中的电子元器件因制造容差或漂移容差等物理因素导致的固有特性, 无法改变和伪造。利用设备指纹对设备进行识别和认证具有难以克隆和伪造以及对上层协议透明等优点, 目前已在无线网络认证中引起了高度关注, 在无线 WiFi、LTE 和 Zigbee 等无线系统都有广泛研究[1-3]。例如, 彭林宁等人[4]利用差分星座轨迹图(CTF)和来自 CTF 的载波频偏、调制偏移与 I/Q 偏移等多个设备指纹特征, 很好地实现了对 Zig-bee 设备的分类识别。

有线网络中, 一张普通的网卡也含有大量的电子元器件, 电子元器件也会因制造容差和漂移容差而造成不同的物理特征。此外, 网卡工作环境(如温度、噪声、传输通道等因素)的变化也会导致其物理特征的变化[5]。文献[6]的研究表明, 这些物理层特征在不同网卡中是唯一的, 即使是同一厂家同一型号同一系列甚至是同一批次的通信设备也会存在微小差异。因此, 受无线系统中设备指纹识别成功的启发, 自 2006 年起, 有线领域也开始了基于设备指纹识别和认证的研究[7-8]。2011 年, Gerdes 首次证实了从以太网网卡信号中可以提取出有效的网卡指纹[9]。后来, 陆续有专家研究有线网络设备指纹的提取和识别, 但截至目前, 国内外涉足有线网络设备指纹的提取与识别的研究仍然较少。

2 相关知识

2.1 设备指纹的提取

无线设备的指纹可从瞬态信号中提取, 也可从稳态信号中提取。瞬态信号[10]是设备开/关瞬间产生的信号, 不包含任何数据信息, 只体现设备的硬件特征, 具有“独立性”, 射频指纹最初就是从瞬态信号中提取的[11]。但是瞬态信号持续时间短, 难以捕获, 而且需要采样率很高的接收机(比如高性能示波器), 对突变点检测和定位较为敏感, 需要高信噪比条件, 这些弊端限制了基于瞬态信号的设备指纹的应用。而且, 根据 Gerdes 等人[12]的研究, 瞬态响应在有线网络的传输过程中非常微弱, 获得的数据不足以为有线网卡的区分提供充足的特征, 且很难在帧头捕捉到瞬态信号。因此, 瞬态信号不适用于有线设备的指纹提取。因而本文主要研究从稳态信号中提取网卡指纹的方法。

稳态信号是指瞬态信号之后, 设备处于稳定工作状态时的信号。稳态信号持续时间长, 更容易获得, 利用廉价的接收机即可完成。目前针对有线设备指纹的提取也都基于稳态信号, 主要的研究有: 2012年, Gerdes[12]利用自适应滤波结合傅里叶变换提取出网卡指纹, 对27个10M 以太网网卡的识别率达到94.0%; 2015~2016年, Carbino 等人[13-14]利用基于星座独特本质属性(CB-DNA)的方法, 将以太网卡无意的电缆辐射信号映射到二维星座空间从而提取出网卡指纹, 对16个10M 以太网网卡的识别率达到93.1%; 2017年, Ross[15]利用有线信号独特本地属性(WS-DNA)方法从电力线通信系统中的 Hub 设备信号中提取指纹, 对6个 Hub 设备在信噪比 SNR=46时识别率达到99.0%; 2019年, 国内的彭林宁等人[16]利用临近星座恍惚状态图(ACTF)提取出光纤以太网设备的指纹, 对24个光纤以太网设备的识别率高达99.49%。

目前, 针对有线网卡指纹提取的研究成果仅有上面这些。这些研究大都基于 10M 有线网卡, 而当前10M有线网卡已非主流, 而是被更高的100M及以上速率的网卡所替代。另外, 研究的样本量较低, 网卡识别率也普遍不高。相对于一块 10M 有线网卡发出的前导码信号固定不变, 为了避免出现连续长度的重复数据, 保证输出均匀功率的数据信号, 100M 有线网卡在发送数据时会对数据进行加扰[17-18], 这样一块 100M 有线网卡发出的前导码信号会随机变动, 因此, 不能使用前导码信号提取特征。另外, 100M 有线网卡不再如 10M 有线网卡那样采用曼彻斯特编码(用跳变沿表示二进制 0 或 1, 从高电平到低电平跳变表示 0, 反之表示 1), 而是先进行 4B/5B 编码再进行三电平的 MLT-3 编码[19-21](信号分正电平、负电平、零电平三种电位状态, 用电平不变表示二进制 0, 用电平跳变表示二进制 1), 这样二进制 0 的波形没有变化, 二进制 1 的波形变化不固定, 基于波形

星座图的特征提取方法不适用。因此, 本文针对 50 块 100M 速率的以太网网卡展开研究, 提出一种基于最小均方误差(Least Mean Square)自适应滤波算法(以下简称 LMS 算法)的指纹提取方法, 该方法能够直接从 100M 以太网网卡的稳态信号中提取出网卡指纹, 无需信号的任何先验知识, 简单方便, 且提取的网卡指纹产生自网卡本身的物理特性, 不可克隆, 无法被篡改。

2.2 LMS 自适应滤波算法

LMS 算法[22]是一种自适应滤波算法, 自适应滤波算法的核心思想就是调整滤波器自身参数, 使滤波器的输出信号与期望输出信号之间满足某种最佳准则要求。不同的准则可以产生不同的自适应滤波算法, 如基于最小均方误差、递推最小二乘(RLS)、变换域、子带分解以及 QR 分解等自适应滤波算法。其中, 由 Widow 和 Hoff 提出的最小均方误差算法(以下简称 LMS 算法), 具有计算量小、易于实现、稳定性好等优点而在实践中被广泛采用。

LMS 算法的准则是使均方误差达到最小, 即期望响应与滤波器实际输出之差的平方的期望值达到最小, 并且依据这个准则来修改权重系数向量。如图 1 所示为 LMS 算法原理图。

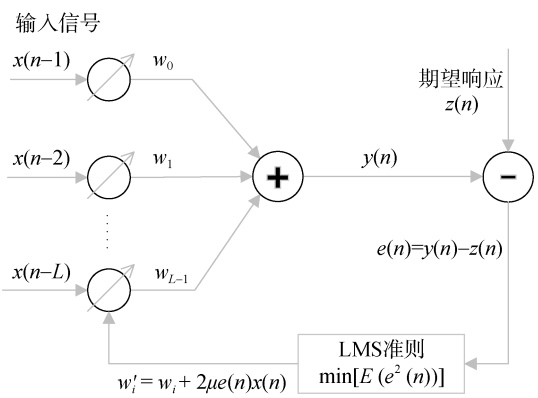


图 1 LMS 算法基本原理

Figure 1 Basic principle of LMS algorithm

如图所示, 假设输入信号为$x(n)$, 经过自适应滤波器的输出为$y(n)$, $w_i \in \{w_0, w_1, \dots, w_{L-1}\}$为自适应滤波器的权重系数(其中 L 为滤波器的阶数), 那么输入信号$x(n)$经过自适应滤波器滤波后产生的输出信号$y(n)$由如下公式(1)计算:

$$y(n) = \sum_{i=0}^{L-1} w_i x(n-i) \tag{1}$$

滤波器的输出信号$y(n)$与期望响应$z(n)$间的误差为$e(n)$, 由如下公式(2)计算:

$$e(n) = y(n) - z(n) \tag{2}$$

现要求期望响应与滤波器实际输出之间满足 LMS 准则, 即求$\min\left[E\left(e^2(n)\right)\right]$, 使期望响应与滤波器实际输出之差的平方的期望值$E(e^2(n))$达到最小。

根据最速下降理论, 沿着目标函数最速下降方向(即负梯度方向)调整滤波器权重系数, 会找到目标函数的最小值, 因而对$\min\left[E\left(e^2(n)\right)\right]$求导并使导数为 0, 得到:

$$w_i' = w_i + 2\mu e(n)x(n) \tag{3}$$

按照公式(3) (μ 为收敛因子)来迭代修改自适应滤波器的权重系数 W, 期望响应与滤波器实际输出之差会越来越小, 滤波器实际输出会无限逼近期望响应。给定收敛因子 μ, 滤波器的权重系数最终会收敛到一个固定值, 此时滤波器实际输出与期望响应之间的误差最小。

LMS 算法可不需要任何关于目标信号的先验知识而可直接从目标信号中提取出其特征指纹。鉴于目前国内外对有线以太网网卡指纹的研究成果较少, 对于有线以太网网卡的信号特征缺少先验知识, 因此本文提出基于 LMS 算法提取有线以太网网卡指纹。

3 网卡信号采集

3.1 采集分析

本文研究从有线以太网网卡的稳态信号中提取网卡的物理特征。如引言所述, 网卡的工作环境如温度、噪声、传输通道(包括传输介质、介质长度)等因素也会引起物理特征的变化, 这些因素为网卡信号增加了随机和非平稳成分, 必须尽可能地将其差异最小化以获得一致的测量结果。为此, 本文尽可能地在同一环境下采集网卡信号, 这样, 信号的传输通道相同, 采集信号时的温度、噪声变化不大, 网卡的物理特征因此而产生的差异可忽略不计。

有线以太网网卡一般作为用户 PC 的网络接口卡与交换机互联, 其网络环境是遵守 IEEE 802.3 标准的以太网[23]。以太网使用的传输介质主要包括同轴电缆、双绞线、光纤。根据不同的传输速率和距离要求, 基于这三类介质的信号线又衍生出很多不同的种类。目前最主流的传输介质是 100BASE-T 和 1000BASE-T 的 5 类双绞线, 它们的传输速率分别为 100Mbps 和 1000Mbps。考虑 1000BASE-T 双绞线即使工作在半双工模式下, 在空闲时交换机也会给网卡传输数据, 这会干扰网卡发出的信号, 因此本文采集基于 100BASE-TX 标准传输(采用 5 类双绞线、100Mbps 传输速率)的网卡信号。

IEEE 802.3 标准规定以太网中传输的数据包格式如图 2 所示, 该数据包格式被称为以太网 MAC 帧。

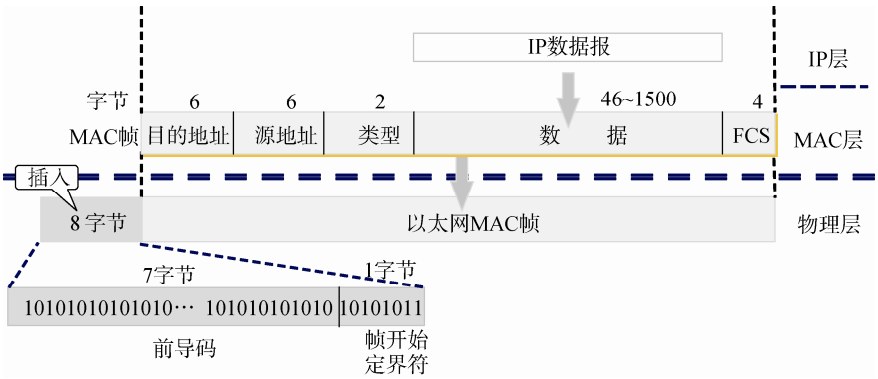


图 2　以太网 MAC 帧

Figure 2　Ethernet MAC frame

在传输每个以太网 MAC 帧时, 网卡会先发一段前导码, 内容为连续 7 个字节的 0x55, 用于使收发双方的时钟同步; 然后发一个字节的帧开始定界符, 内容为 0xD5, 用于指示以太网 MAC 帧的开始。随后的以太网 MAC 帧的每个字段会随传输数据和网卡的不同而变化。如 2.1 节所述, 100M 有线网卡在发送数据时会对数据进行加扰[33], 导致发出的前导码信号会随机变动, 不能使用 100M 有线网卡的前导码信号提取特征。因此, 本文采集包含前导码在内的若干个完整的以太网 MAC 帧信号, 从长信号段中提取指纹特征以降低扰码对信号波形的影响。同时, 为了尽量减少数据对网卡信号的影响, 本文采集网卡空闲时的信号, 也就是不传输用户数据时的信号。

3.2　采集实施

在有线网卡发送数据信息时, 传输介质上便会产生相应的模拟电平信号, 这些信号按照 IEEE 802.3 标准约定的数据传输协议与数据编码协议生成, 需要在保留其物理层特征的前提下进行采集, 作为提取网卡指纹的原料。为此, 本文采用过采样方式在网卡和交换机之间的线路上采集以太网网卡的完整的两路差分信号。采集网卡信号的组网如图 3 所示, 待采集的网卡通过 100Mbps 速率的双绞线连接交换机, AD9484 信号采集板串联在网卡与交换机之间, 使用

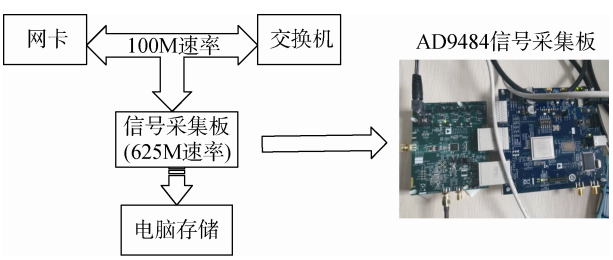


图 3　网卡信号采集组网

Figure 3　The network for network Ethernet card signal collection

625Mbps 的采样率、8bit/样点的采样精度来采集网卡与交换机间的闲时信号(即不传输用户数据情况下的信号), 采集的信号经过 USB 接口存储到电脑中。

本次实验实际采集到的网卡信号如图 4 所示。而 100BASE-TX 标准传输的网卡信号按 MLT-3 编码方式编码[32], 其标准信号波形如图 5 所示。

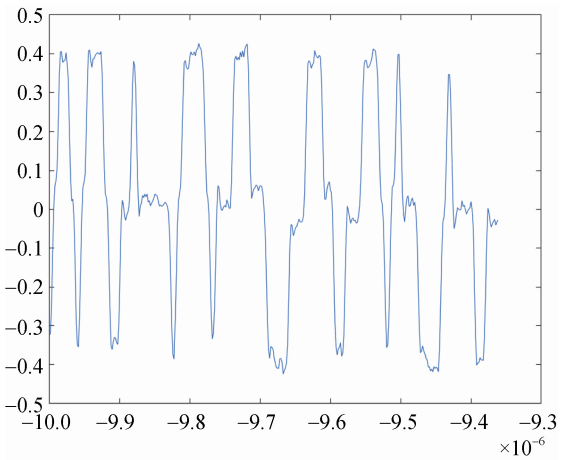


图 4　采集到的网卡差分信号

Figure 4　The collected differential signal of Ethernet card

对比图 4 和图 5 可以发现, 实际的网卡信号波形与标准的 MLT-3 编码波形相比, 并没有那么完美, 存在着许多抖动与毛刺, 而这其中正包含着网卡的物

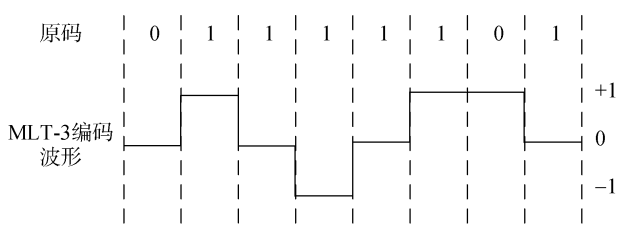


图 5　MLT-3 编码的标准信号波形

Figure 5　Standard signal waveform encoded by MLT-3

理特征。从采集到的网卡信号中将这些特征提取并数据化, 就可以作为发送设备的身份认证信息。在本文中, 将提取的这些特征信息统一称为“网卡指纹”。

4 网卡指纹提取系统

本文利用 LMS 算法提取有线网卡指纹, 如图 6 所示为提取指纹的整个过程: 首先将采集的网卡信号进行预处理, 以消除影响指纹提取的因素; 再通过判决整形和符号同步获得与接收信号同步的本地时钟信号, 用此同步时钟对判决整形后的信号重新采样, 即可获得与接收信号同步的理想信号。将理想信号作为 LMS 自适应滤波器的滤波输入, 预处理后的接收信号作为 LMS 自适应滤波器的期望响应, 使得理想信号在滤波后以最小均方误差逼近接收信号。自适应滤波器收敛后, 采用滤波器的权重系数作为网卡指纹。

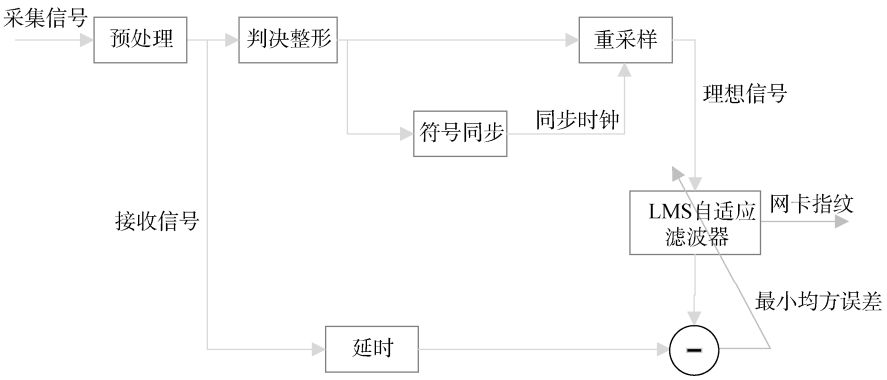


图 6 网卡指纹提取系统

Figure 6 The system of Ethernet card fingerprint extraction

用上述方法提取同一网卡多个采集信号的指纹信息, 取这些指纹信息的平均值, 即可获得较为稳定的网卡指纹。

4.1 预处理

预处理主要目的是消除环境因素对采集信号的影响, 以方便特征提取。一般包括直流分量、功率归一化等。但本文采集的网卡信号采用曼彻斯特编码, 该编码下每位编码中信号都会跳变一次, 因此不存在直流分量, 具有良好的抗干扰性能, 因而无需去除直流分量。但网卡和交换机之间距离的变化会导致接收功率的变化, 因此需要通过功率归一化来消除接收功率等与采集环境相关的差异。功率归一化方法比较简单。就是首先计算采集信号的各电平幅度的平均值, 然后将各电平除以该平均幅度即可将功率归一化。

为了有效地抑制具体数据对于指纹提取的影响, 信号进行功率归一化处理后, 还需将两路差分信号相减, 使用相减后的信号作为指纹提取的接收信号。

4.2 判决整形和符号同步

如图 6 所示, 在利用 LMS 算法提取网卡指纹系统中, LMS 自适应滤波器需要一个理想信号作为滤波输入, 该理想信号须与接收信号同步, 否则两个信号段的起始位置在一个周期内可能是不同的, 这样会弱化提取特征的有效性。本文设计先通过“判决整形”和“符号同步”获得与接收信号同步的本地时钟信号, 再使用该本地时钟信号重新采样判决整形后的信号, 即可获得与接收信号同步的理想信号。

4.2.1 判决整形

判决整形的目的是将接收信号的波形整形为与理想的差分曼彻斯特编码波形相似的方波, 本文将判决门限设定为接收信号各个电平的平均幅度值的 1/2。如图 7 所示为接收信号判决整形后的波形与接收信号波形对比图。

4.2.2 符号同步

符号同步的目的是获得与接收信号同步的本地时钟信号。为此, 本文通过提取判决整形后的信号的上升沿和下降沿生成脉冲序列, 并在本地生成一个与信号数据速率相同的本地时钟信号, 然后通过循环移位并计算脉冲序列与本地时钟信号的异或的和, 异或和最小的本地时钟信号即为与接收信号同步的本地时钟信号。

如图 8 所示为获取的理想信号波形与接收信号波形的对比图。从图中可以看到, 理想信号与接收信号的时钟是同步的。

4.3 利用 LMS 算法提取网卡指纹

如图 6 所示, 将上面获得的理想信号作为 LMS 自适应滤波器的输入信号, 将预处理后的接收信号经过延时后作为期望响应, 分别送入 LMS 自适应滤

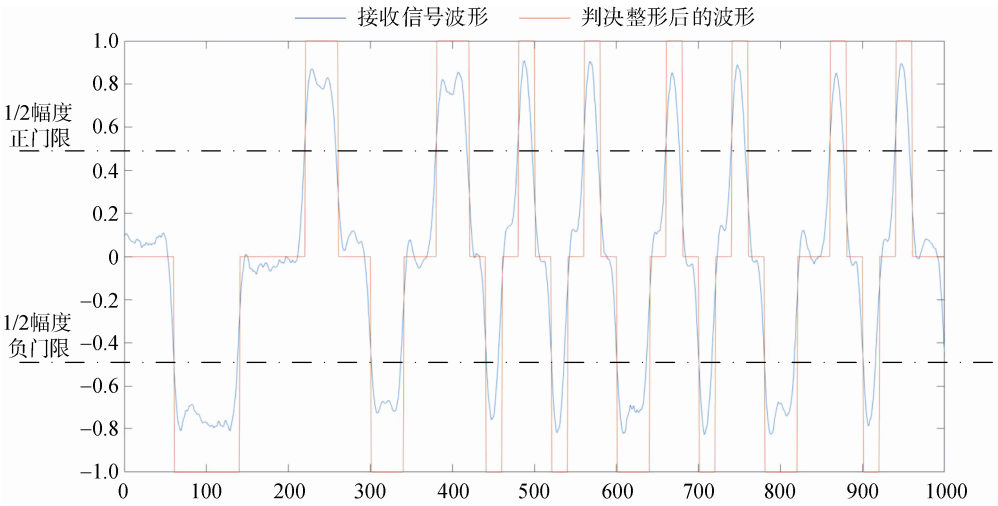


图 7　判决整形效果图

Figure 7　The result diagram of the decision shaping

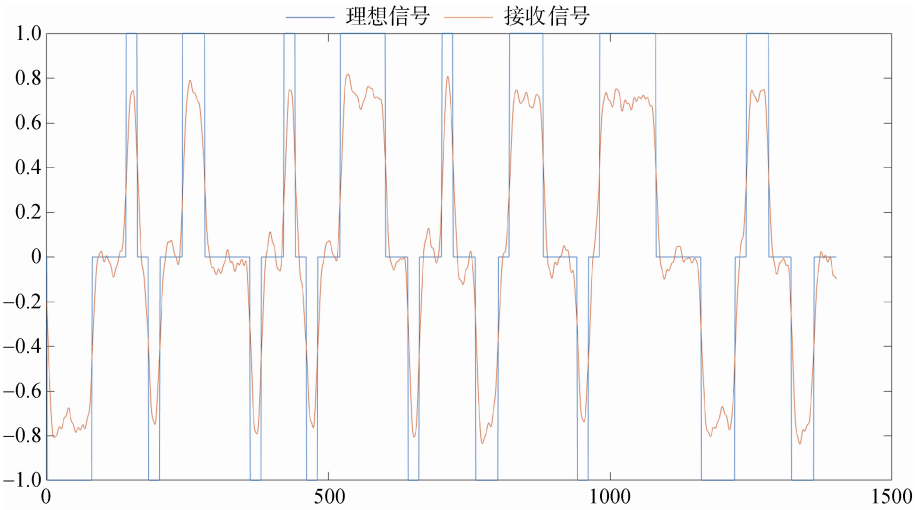


图 8　理想信号波形与接收信号波形比对

Figure 8　Comparison between the ideal signal waveform and the actual received signal waveform

波器, 使得理想信号在滤波后以最小均方误差逼近接收信号。

假设接收信号包含 M 个数据, 接收信号为 $\{z_m(n), m = 0,1,\dots,M-1\}$, 经判决整形和重采样后获得的理想信号为$\{x_m(n), m = 0,1,\dots,M-1\}$, 则理想信号经 LMS 自适应滤波器滤波后产生的输出$y_m(n)$为:

$$y_m(n) = \sum_{i=0}^{L-1} w_i x(n-i) \tag{4}$$

$y_m(n)$与接收信号$z_m(n)$间的误差为$e(n)$, 即:

$$e(n) = \{z_m(n)\} - \{y_m(n)\} \tag{5}$$

然后使用$w_i' = w_i + 2\mu e(n)x_m(n)$(其中 μ 为滤波器收敛因子)迭代调整自适应滤波器的权重系数$w_i \in \{w_0, w_1, \dots, w_{L-1}\}$(其中$L$为滤波器阶数)至滤波器收敛。

LMS 自适应滤波器收敛后, 其输出信号与接收信号的最小均方误差达到最小, 此时认为理想信号在滤波后与接收信号基本相同, 而自适应滤波器的权重系数$w_i \in \{w_0, w_1, \dots, w_{L-1}\}$(其中 L 为滤波器阶数)则认为是接收信号与理想信号之间的差异特征, 该差异特征是不同网卡的硬件特性对网卡信号的影响而产生的, 具有唯一性, 可作为网卡指纹。

通过上述方法对 K 个包含 M 个数据的接收信号进行指纹提取, 取 K 组自适应滤波器的权重系数的平均值作为网卡指纹输出, 即可获得较为稳定的网卡指纹, 计算公式如下所述。

$$\bar{W} = \frac{1}{K}\left\{\sum_{i=1}^{K} w_{0,i}, \sum_{i=1}^{K} w_{1,i}, \dots, \sum_{i=1}^{K} w_{L-1,i}\right\} \tag{6}$$

5　实验结果

5.1　数据集

本文实验用到的数据集来自 6 个品牌共 50 块有

线网卡, 每块网卡不同类型。设备品牌和索引分别为锐捷(索引 1~9)、绿联(索引 10~29)、小米(30~34)、山泽(35~42)、TP-LINK(43~46)和 CE-LINK(47~50)。对 50 块网卡信号进行过采样(采样率为 625MHz、采样精度为 8 比特/样点), 每块采样 1 次共得到 100 个波形段(每次差分采样产生 2 个波形段, 使用时合成 1 个波形段), 每个波形段捕获 625 万个样本点。再将每个波形段的 625 万个样本按照设定的数据长度划分成若干个信号片段(假设数据长度 10000, 则每个波形段划分为 600 个信号段), 形成数据集 1, 其中 1/3 的信号段用作训练集, 2/3 的信号段用作验证集。

5.2 滤波参数确定

LMS 算法的主要参数有数据长度 L、滤波器阶数 N、收敛因子 μ、迭代次数 K($N<K\leqslant L$, 本文实验中统一选择与数据长度相等)。

为了确定数据长度 L、滤波器阶数 N、收敛因子 μ 三个参数相对较优的参数设置, 本文通过先变化一个参数、固定其他参数, 然后依据收敛性、收敛速度、稳态误差和计算复杂度等主要性能指标来确定相对较优的参数设置, 采用数据集 1 进行参数确定实验。

5.2.1 数据长度确定

选取滤波器阶数 N=261、收敛因子 μ=0.0001、迭代次数 K=L, 然后观察滤波器系数在达到最优值前的理想信号与实际信号之间的误差变化曲线, 该曲线反映了滤波器权重系数的收敛性。实验中, 取初始数据长度 L=100、500、2000、4000、8000、16000、50000、100000 和 200000, 分别用这些数据长度提取网卡指纹, 并记录滤波器收敛过程中的误差值, 误差值随数据长度 L 变化如图 9 所示。

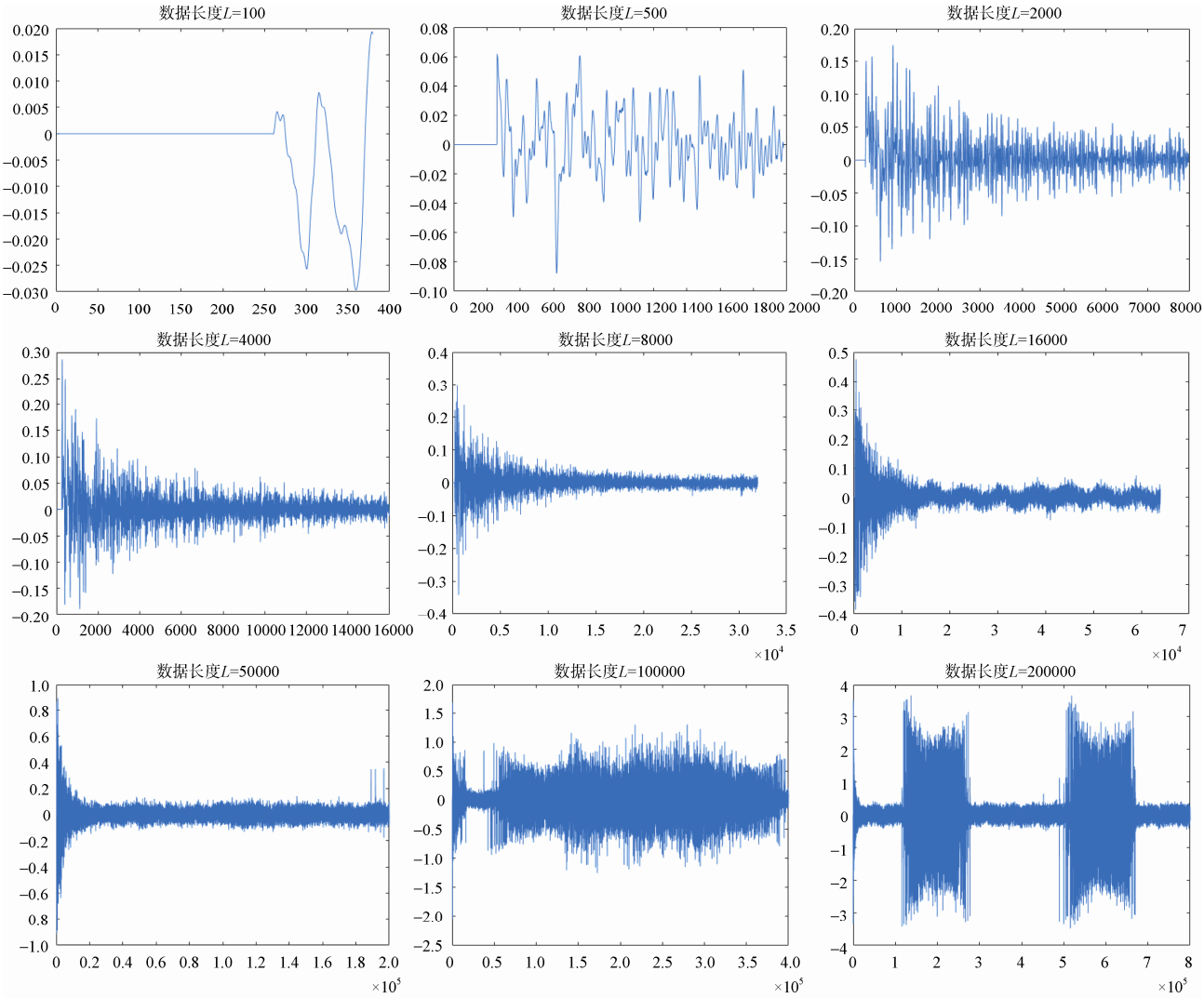


图 9 误差随数据长度变化曲线

Figure 9 The variation curve of error with data length

从图 9 可以看到, 理想信号与实际信号之间的误差的震荡幅度随着数据长度的增加而增大。数据长度 L 低于 100 时误差曲线不收敛; 超过 16000 后, 误差曲线呈现震荡收敛; 超过 50000 后, 误差曲线不再单向收敛。因此, 滤波器权重系数在数据长度 100<L<16000 时, 具有较好的收敛性, 而在高于 50000 时, 不再单向收敛。同时, 根据上述公式(5), 数据长度越大, 计算量越大, 收敛速度也就越慢。因此, 综合考虑收敛性和收敛速度, 数据长度推荐选取在[100,16000]之间。

5.2.2 滤波器阶数确定

选取 L=10000、μ=0.0001、K=10000, 分别按照滤波器阶数 N=[5,25,45,65,85,…, 985, 1005]提取网卡指纹, 然后计算使用每一个滤波器阶数提取指纹时到达平稳状态后的误差(即稳态误差)的幅度平均值, 观察该平均值随滤波器阶数 N 的变化, 如图 10 所示。

从图 10 可以看出, 在滤波器阶数为 5~165 范围内时, 稳态误差幅度的平均值随着滤波器阶数的增加而减小, 在滤波器阶数高于 165 时, 滤波器阶数的增加对稳态误差幅度的平均值影响不大, 阶数高于 285 时, 稳态误差幅度的平均值基本不再变化。

考虑滤波器阶数越大, 计算复杂度越高, 因此, 滤波器阶数推荐设置在[165,285]之间。

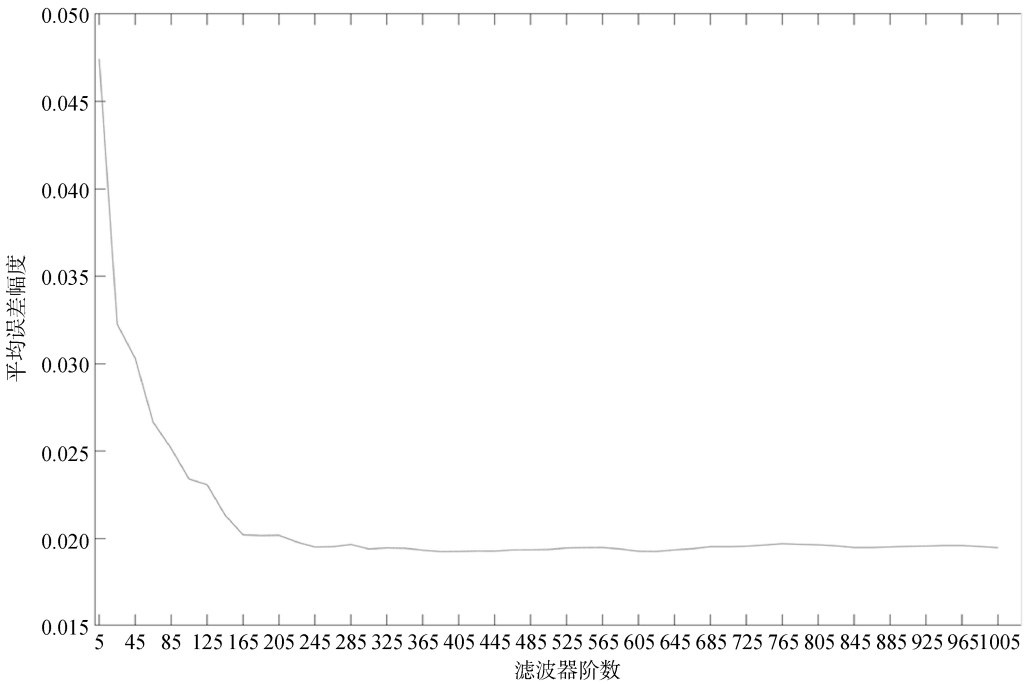


图 10 误差幅度随滤波器阶数变化曲线

Figure 10 The variation curve of error amplitude with filter order

5.2.3 收敛因子确定

1996 年, Hayjin 证明, 只要收敛因子满足下式, LMS 算法就是按方差收敛的。

$$0 < \mu < \frac{2}{\lambda_m} \tag{7}$$

其中, λ_m 是输入向量$x_m(n)$(即接收信号)组成的自相关矩阵 $\boldsymbol{R}$ 的最大特征值。由于 λ_m 常常不可知, 因此, 往往使用自相关矩阵 $\boldsymbol{R}$ 的迹来代替。按定义, 矩阵的迹是矩阵主对角线元素之和:

$$\mathrm{tr}(R) = \sum_{i=1}^{Q} R(i,i) \tag{8}$$

同时, 矩阵的迹又等于矩阵所有特征值之和, 因此一般有 tr(R)>λ_m。只要取:

$$0 < \mu < \frac{2}{\mathrm{tr}(R)} < \frac{2}{\lambda_m} \tag{9}$$

即可满足收敛条件。按定义, 自相关矩阵的主对角元素就是各输入向量的均方值。因此公式又可写为:

$$0 < \mu < \frac{2}{x(n)\text{向量均方值之和}} \tag{10}$$

按此公式计算, 收敛因子的取值范围大概为 0<μ<0.2。

本次实验中, 分别选取 μ=0.1, μ=0.01, μ=0.001, μ=0.0001, μ=0.00001, 并设置数据长度 L=10000, 滤波器阶数 N=245, 迭代次数 K=10000 来提取网卡指纹, 到达平稳状态后, 理想信号与实际接收信号的误差随收敛因子 μ 的变化如图 11 所示。

从图 11 可以看出, 当 μ>0.01 时误差不收敛, 因此滤波器权重系数也不具备收敛性。当 0<μ<0.001 时, 误差收敛, 因而自适应滤波器收敛, 但是, 随着

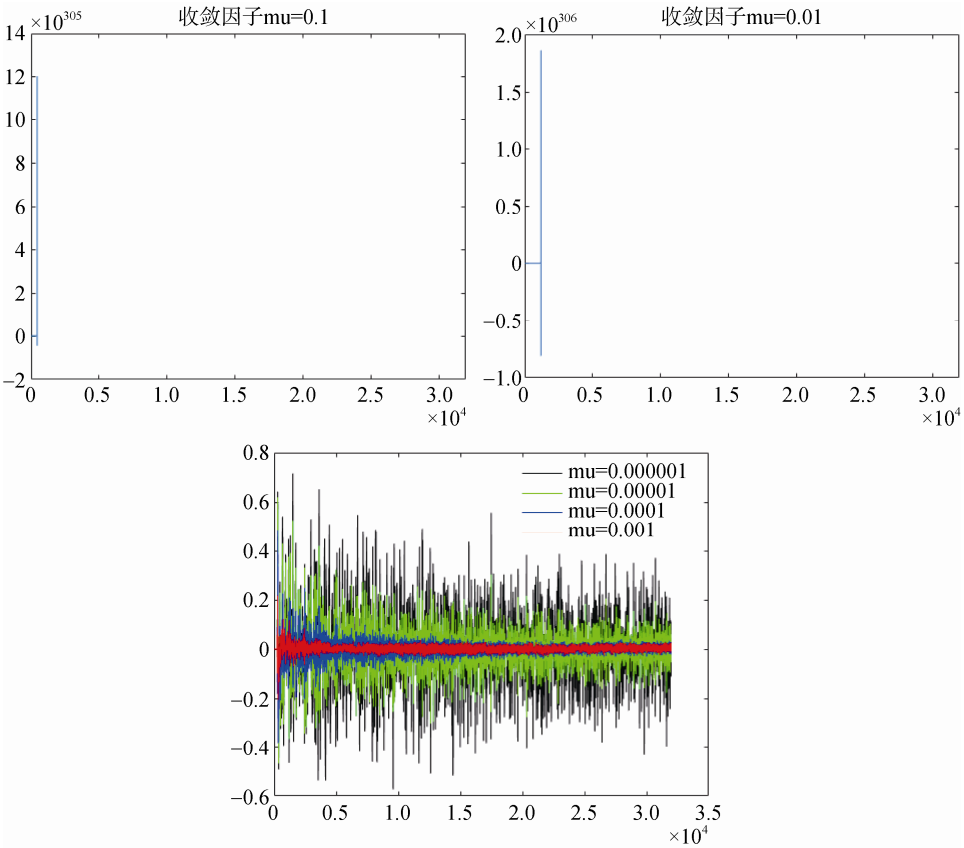


图 11　误差随收敛因子变化图

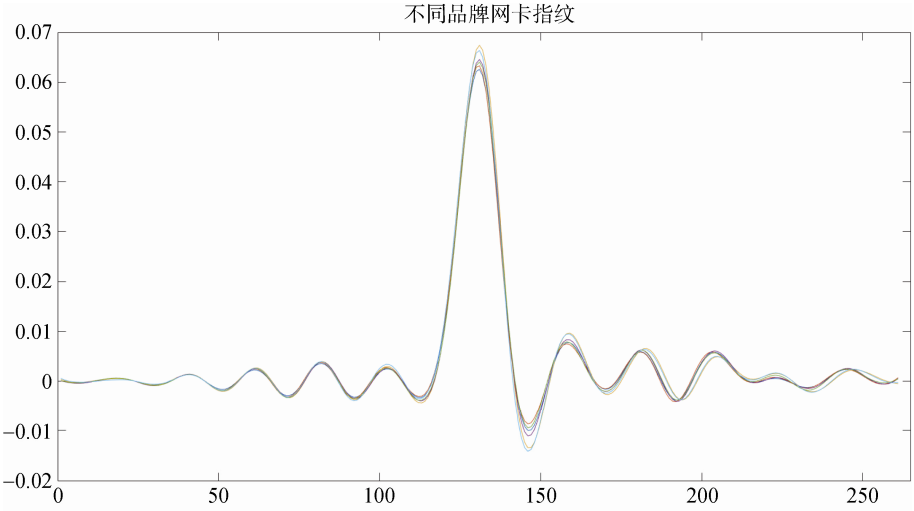
Figure 11　The variation graph of error with convergence factor

收敛因子减小, 误差收敛速度变慢, 误差震荡幅度变大。因此, 综合收敛速度和收敛性, 收敛因子推荐选取在[0.00001,0.001]之间。

5.3　提取网卡指纹

设置数据长度 L=10000、收敛因子 μ=0.0001、滤波器阶数 N=261, 迭代次数 K=L, 使用数据集 1 提取指纹, 然后分别选取不同品牌网卡的指纹、同一品牌不同网卡的指纹进行对比分析, 结果分别如图 12 所示。

从图中可以看出, 无论同品牌, 还是不同品牌, 均能从有线网卡信号中提取出唯一且相对稳定的特征指纹, 但不同网卡之间的指纹差异性比较微弱, 区分性小, 尤其是同一品牌不同类型网卡之间的指纹差异性非常微弱。这给指纹的提取和分类识别带来困难。



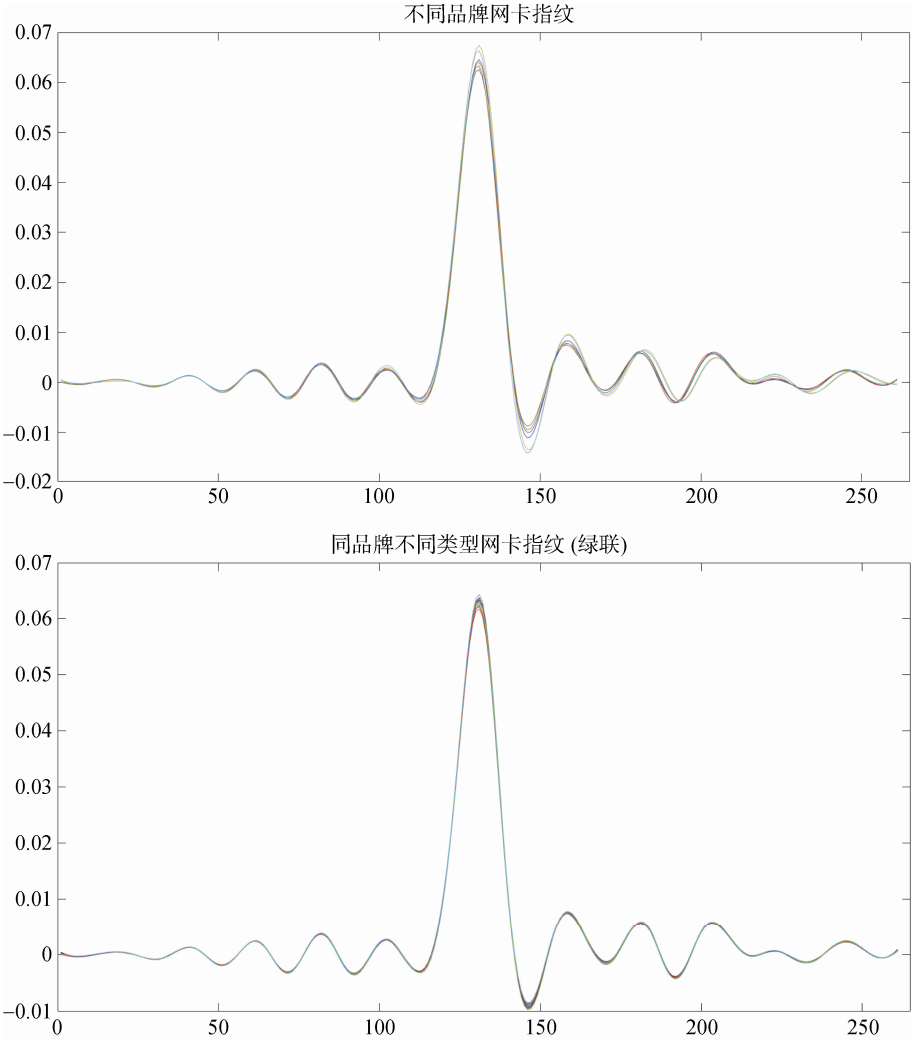


图 12　不同品牌网卡和同品牌不同类型网卡的指纹

Figure 12　Fingerprints of Ethernet cards of different brands and different types of Ethernet cards of the same brand

5.4　指纹有效性验证

为了验证提取的网卡指纹的有效性, 本文使用 Matlab 自带的分类器对数据集 1 和数据集 2 提取出的网卡指纹分别进行了分类识别。

5.4.1　分类算法选取

Matlab 自带有多种分类器, 包括决策树、判别分析、支持向量机 SVM、最近邻分类器、朴素贝叶斯分类器、集成分类器, 每一种分类器又包含多种算法。为了确定选取哪些分类算法较优, 本文先使用 Matlab 自带的所有分类器对训练样本进行分类训练, 交叉验证折数分别设置为 5 折、10 折和 15 折, 然后获取训练结束后的分类准确率, 选取准确率较高的分类算法。通过实验, 当交叉验证折数为 10 折时, 分类准确率普遍较高, 较高的 3 个分类算法分别为线性判别、线性 SVM 和集成子空间判别。因此, 后续使用线性判别和集成子空间判别这 3 种算法对网卡进行分类识别, 且交叉验证折数设置为 10 折。

5.4.2　使用网卡指纹分类识别

将数据集中的每个波形段划分为 600 个信号片段, 其中前 200 个信号段用于训练, 后 400 个信号段用于验证。提取每个信号片段的特征指纹, 并取 20 个信号片段的特征指纹的平均值作为该波形段对应网卡的网卡指纹, 用每个波形段前 200 个信号段提取出的 10 个网卡指纹进行分类器训练, 后 400 个信号段提取出的 20 个网卡指纹进行分类识别验证。

设置 μ=0.0001, N=261, L=2000、5000、10000、15000、20000, 并设置迭代次数 K=L, 交叉验证折数设置为 10 折, 网卡识别率如表 1 所示。

设置 N=261, L=5000, μ=0.00001、0.00005、0.0001、0.0005、0.001, 并设置迭代次数 K=L, 交叉验证折数设置为 10 折, 网卡识别率统计如表 2 所示。

设置 μ=0.0001, L=5000, N=21、81、165、261、501、1001, 并设置迭代次数 K=L, 交叉验证折数设置为 10 折, 网卡识别率统计如表 3 所示。

表 1　网卡识别率随数据长度变化(μ=0.0001, N=261)

Table 1　Change of Ethernet card recognition rate with data length(μ=0.0001, N=261)

分类算法/数据长度	L=100	L=500	L=1000	L=2000	L=5000	L=10000	L=15000	L=20000
线性判别	0.971	0.999	0.999	0.998	0.999	0.997	0.997	0.991
线性 SVM	0.474	0.982	0.992	0.996	0.990	0.962	0.953	0.920
集成子空间判别	0.973	0.999	0.999	1	1	0.999	0.999	0.998

表 2　网卡识别率随收敛因子变化(N=261, L=5000)

Table 2　Change of Ethernet card recognition rate with convergence factor (N=261, L=5000)

分类算法/收敛因子	μ=0.001	μ=0.0005	μ=0.0001	μ=0.00005	μ=0.00001
线性判别	0.973	0.990	0.999	0.999	1
线性 SVM	0.913	0.955	0.990	0.992	0.997
集成子空间判别	0.985	0.990	1	1	1

表 3　网卡识别率随滤波器阶数变化(μ=0.0001, L=5000)

Table 3　Change of Ethernet card recognition rate with filter order (μ=0.0001, L=5000)

分类算法/滤波器阶数	N=21	N=41	N=81	N=165	N=261	N=501	N=1001
线性判别	0.997	0.992	1	1	0.999	0.998	0.990
线性 SVM	0.954	0.969	0.992	0.994	0.990	0.978	0.969
集成子空间判别	0.999	0.990	1	1	1	1	0.998

从表 1 可以看出, 在收敛因子和滤波器阶数固定情况下, 网卡识别率随数据长度先增加后减少, 在 L=[1000,5000]时识别率较高; 从表 2 可以看出, 在数据长度和滤波器阶数固定情况下, 收敛因子越小, 网卡识别率越大, 但同时收敛因子越小, 收敛时间越长; 从表 3 可以看出, 在数据长度和收敛因子固定情况下, 网卡识别率在滤波器阶数处于 81~261 之间时较高。同时从表 1、表 2 和表 3 也可看出, 使用线性判别和集成子空间判别两种分类算法时, 本文方法提取指纹的网卡识别率较高, 前者可在 97.1%以上, 后者可在 97.3%以上。

如图 13 所示为线性识别和集成子空间判别两种分类算法在识别率为 0.973 和 0.985 时的混淆矩阵, 其中横坐标为识别出的网卡索引, 纵坐标为实际网卡索引。

从图 13 中可以看出, 使用线性判别分类算法时, 错误识别的网卡索引对主要包括: 12->15(即网卡 12 错误识别为网卡 15)、10->17、13->20、21->23/25、22->21、27->20、28->13、29->10、40->44; 使用集成子空间判别分类算法时, 错误识别的网卡索引对主要包括: 12->15、13->20、16->18、17->16/29、20->13、21->22、22->21、26->27、29->10。

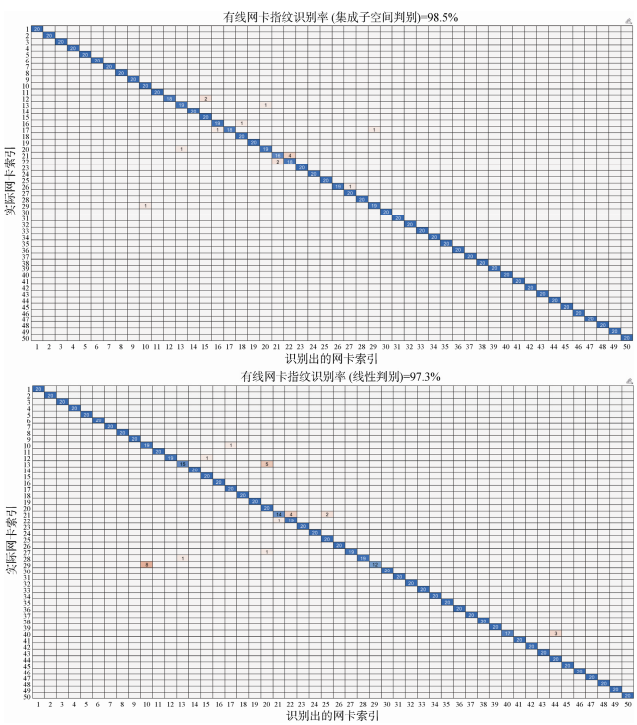


图 13　网卡指纹识别混淆矩阵

Figure 13　The confusion matrix for the identification of Ethernet card by NIC fingerprints

根据前文所述, 索引 10~29 对应绿联网卡, 索引

40 对应山泽网卡, 索引 44 对应 TP-LINK 网卡, 因此使用本文方法提取的网卡指纹识别网卡时, 错误主要发生在同品牌的绿联网卡中; 不同品牌的指纹识别只将索引为 40 的山泽网卡错误识别为索引为 44 的 TP-LINK 网卡, 识别效果比同品牌不同类型网卡好。其根本原因如 5.3 章节所述, 不同品牌网卡的指纹之间的差异较大些, 而同一品牌不同类型网卡的指纹差异较小。

6 总结

本文提出了一种基于 LMS 自适应滤波算法的有线网卡指纹提取方法, 该方法不需要有线网卡信号特征的先验知识, 可直接从 100M 以太网网卡信号中提取出网卡指纹。通过实验证明, 有线网卡确实存在特征指纹, 但不同网卡指纹区分性小, 当数据长度选取在 1000~5000 范围内、滤波器阶数选取在 81~261 范围内、收敛因子选择在 0.00001~0.0001 之间时, 使用本文方法可有效提取出有线网卡指纹, 且可获得较好的识别效果(考虑收敛性、收敛速度、计算复杂度和稳定误差等因素)。

然后本文使用 Matlab 自带的分类器对提取的有线网卡指纹进行训练, 利用训练的模型对网卡进行分类识别。十折交叉验证结果表明, 使用线性判别和集成子空间判别分类算法时, 网卡识别率可分别达到 97.2%、98.5%以上, 错误的指纹识别主要发生在同品牌网卡中, 不同品牌的网卡指纹识别效果较好。

利用本文方法提取网卡指纹, 简单方便, 且提取的网卡指纹产生自网卡本身的物理特性, 不可克隆, 无法被篡改, 利用这样的指纹再结合合适的分类学习算法就可以对以太网网卡进行有效识别, 进而可方便且可靠地实现对有线网终端设备的接入认证。在大规模控制和自动化通信领域(如电网)、光纤通信领域, 以及金融、保险、公安等有线专网领域, 目前迫切需要轻量级且安全可靠的接入认证方法, 本文提出的方法可以预见地具有广阔的应用前景。

参考文献

[1] Vo-Huu T D, Vo-Huu T D, Noubir G. Fingerprinting Wi-Fi Devices Using Software Defined Radios[C]. *The 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, 2016: 3-14.

[2] Demers F, St-Hilaire M. Radiometric Identification of LTE Transmitters[C]. *2013 IEEE Global Communications Conference*, 2013: 4116-4121.

[3] Rondeau C M, Betances J A, Temple M A. Securing ZigBee Commercial Communications Using Constellation Based Distinct Native Attribute Fingerprinting[J]. *Security and Communication Networks*, 2018, 2018: 1489347.

[4] Peng L N, Hu A Q, Zhang J Q, et al. Design of a Hybrid RF Fingerprint Extraction and Device Classification Scheme[J]. *IEEE Internet of Things Journal*, 2019, 6(1): 349-360.

[5] Yu J B, Hu A Q, Zhu C M, et al. RF Fingerprinting Extraction and Identification of Wireless Communication Devices[J]. *Journal of Cryptologic Research*, 2016, 3(5): 433-446.
(俞佳宝, 胡爱群, 朱长明, 等. 无线通信设备的射频指纹提取与识别方法[J]. *密码学报*, 2016, 3(5): 433-446.)

[6] Yuan H L, Hu A Q. Fountainhead and Uniqueness of RF Fingerprint[J]. *Journal of Southeast University* (*Natural Science Edition*), 2009, 39(2): 230-233.
(袁红林, 胡爱群. 射频指纹的产生机理与惟一性[J]. *东南大学学报(自然科学版)*, 2009, 39(2): 230-233.)

[7] Daniels T, Mina M, Russell S F. Short Paper: A Signal Fingerprinting Paradigm for General Physical Layer and Sensor Network Security and Assurance[C]. *First International Conference on Security and Privacy for Emerging Areas in Communications Networks*, 2005: 219-221.

[8] Erbskorn J W. Detection of Intrusions at Layer One: A Preliminary Performance Analysis of the IEEE 802.3 Normal Link Pulse as a Means of Host-to-Network Authentication and a Survey of Environmental Effects[D]. Iowa State University, Master of Science, 2009.

[9] Gerdes R. Physical Layer Identification: Methodology, Security, and Origin of Variation[D]. Iowa State University, Doctor of Philosophy, 2011.

[10] Ureten O, Serinken N. Detection of Radio Transmitter Turn-on Transients[J]. *Electronics Letters*, 1999, 35(23): 1996.

[11] Toonstra J, Kinsner W. Transient Analysis and Genetic Algorithms for Classification[C]. *IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings*, 1995: 432-437.

[12] Gerdes R M, Mina M N, Russell S F, et al. Physical-Layer Identification of Wired Ethernet Devices[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(4): 1339-1353.

[13] Carbino T J, Temple M A, Bihl T J. Ethernet Card Discrimination Using Unintentional Cable Emissions and Constellation-Based Fingerprinting[C]. *2015 International Conference on Computing, Networking and Communications*, 2015: 369-373.

[14] Carbino T J, Temple M A, Lopez J. Conditional Constellation Based-Distinct Native Attribute (CB-DNA) Fingerprinting for Network Device Authentication[C]. *2016 IEEE International Conference on Communications*, 2016: 1-6.

[15] Ross B P, Carbino T J, Stone S J. Physical-Layer Discrimination of Power Line Communications[C]. *2017 International Conference on Computing, Networking and Communications*, 2017: 341-345.

[16] Peng L N, Hu A Q. A Design of Deep Learning Based Optical Fiber Ethernet Device Fingerprint Identification System[C]. *ICC 2019 - 2019 IEEE International Conference on Communications*, 2019: 1-6.

[17] Kudinov A, Antimirov Y, Tyshchenko I, et al. The Implementation of the Parallel Scrambler Scheme for the IEEE 802.11 Standard[J]. *International Journal of Electronics & Telecommunications*, 2018, 64(1): 91-94.

[18] Cluzeau M. Reconstruction of a Linear Scrambler[J]. *IEEE Trans-*

actions on Computers, 2007, 56(9): 1283-1291.

[19] Li T. *100BASE-TX Ethernet PHY Chip Design and Verification*[D]. Xi'an: Xidian University, 2018.
(李涛. 基于 100BASE-TX 以太网 PHY 芯片设计与验证[D]. 西安: 西安电子科技大学, 2018.)

[20] Lenell J K. Apparatus for Ethernet PHY/MAC Communication: US8072995[P]. 2011-12-06.

[21] Xu J, Li H L, Zhang F, et al. The Analysis and Implementation of re-Modulated WDM-PON System Based on 4B/5B Coding[J]. *Electronic Design Engineering*, 2017, 25(7): 95-98.
(徐进, 李慧林, 张封, 等. 4B/5B编码的再调制WDM-PON系统分析与实现[J]. *电子设计工程*, 2017, 25(7): 95-98.)

[22] Simon Haykin. Adaptive Filter Theory, Fifth Edition[M]. Beijing: Publishing House of Electronics Industry, 2016: 213.

[23] Kurose J, Ross K. Computer Networks: Top: down Approach[M]. Beijing: China Machine Press, 2009: 931.
(Kurose J, Ross K. 计算机网络: 自顶向下方法[M]. 北京: 机械工业出版社, 2009: 931.)

胡园园 于 2004 年在东南大学生物医学工程专业获得硕士学位。曾在华为从事网络研发工作十多年, 现在东南大学网络空间安全专业攻读博士学位。研究领域为物理层安全。研究兴趣包括: 身份认证、接入控制。Email: 230208463@seu.edu.cn

李晟 于 2019 年在西南交通大学通信工程专业获得学士学位。现在东南大学网络空间安全专业攻读硕士学位。研究领域为物理层安全。研究兴趣包括: 特征工程、数字信号处理。Email: 220194591@seu.edu.cn

李冰 于 2004 年在东南大学微电子与固体电子学专业获得博士学位。现为东南大学微电子学院&网络空间安全学院教授/博导。主要研究方向是高效安全的信息集成电路与系统, 领域包括数字集成电路、嵌入式芯片系统、网络芯片安全与系统。Email: bernie_seu@seu.edu.cn

胡爱群 于 1992 年在东南大学信号与信息处理专业获得博士学位, 现为东南大学信息科学与工程学院教授/博导, 主要研究领域为通信安全、无线网络安全。Email: aqhu@seu.edu.cn

刘佳琦 于 2019 年在东南大学信息工程专业获得学士学位。现在东南大学网络空间安全专业攻读硕士学位。研究领域为物理层安全。研究兴趣包括: 特征提取、身份认证。Email: 220194680@seu.edu.cn