基于WiFi接收信号强度监测的设备类型识别机制*

姜明星[1]，王 玺[2]，郭忠文[2]，王进新[2]
(1.中国海洋大学基础教学中心，山东 青岛 266100;2.中国海洋大学信息科学与工程学院，山东 青岛 266100)

摘 要： 当今WiFi技术迅速发展，WiFi设备的种类和数量随之急剧增长，但WiFi设备识别这方面的研究并不多，仅有的一些研究也过多的依靠beacon节点主动搜集无线信号的方式开展。本文基于非侵入式监测搜集的WiFi信号数据，提出了一种WiFi设备类型的识别机制。通过监测WiFi通信过程中的技术参数，如接收信号强度(Received Signal Strength，RSS)、MAC地址(Media Access Control Address)、通信时间戳等，我们分析提取各类WiFi设备的特征，构建特征向量，然后运用机器学习中相对成熟的分类算法实现对常见无线设备如手机、笔记本电脑和无线路由器的分类。实验结果表明，本文提出的设备类型识别机制使用不同分类算法进行层次化分类后，均可达到较好的效果。

关键词： WiFi监测；接收信号强度；机器学习；层次化分类
中图法分类号： TP391 **文献标志码：** A **文章编号：** 1672-5174(2020)10-134-06
DOI： 10.16441/j.cnki.hdxb.20180167
引用格式： 姜明星，王 玺，郭忠文，等. 基于WiFi接收信号强度监测的设备类型识别机制[J]. 中国海洋大学学报(自然科学版)，2020，50(10)：134-139.
JIANG Ming-Xing，WANG Xi，GUO Zhong-Wen，et al. Identification mechanism of device type based on Wi-Fi signal strength monitoring[J]. Periodical of Ocean University of China，2020，50(10)：134-139.

随着以IEEE802.11为代表的无线通信技术的发展，WiFi相关的无线设备逐步渗透到人类社会的每个角落，生活中无线设备的种类和数量也随之迅速增长。这些无线设备只要处于开机状态并且打开了WiFi开关，就会一直发送WiFi数据包，我们日常生活所处的环境里充斥着无数这样的数据包。与之而来的一个问题就是日益增长的无线网络流量，无线设备的网络流量将在2019年超过有线设备，占据所有IP流量的66%，这一数字比2013年几乎翻了一倍，彼时只有33%的网络流量来自无线设备[1]。

基于这些原因，学术界对无线设备尤其是WiFi设备的研究热情从未消减，关于WiFi设备网络流量的跟踪和分析更是引起了众多研究者的兴趣。同时由于WiFi设备种类繁多，不同类型设备的技术参数和物理特性各不相同：如设备大小、是否有稳定电源、电池容量、操作系统等，这导致它们对应的网络流量也不尽相同，因此对网络流量模型的分析要基于特定的设备类型才有意义[2]。了解WiFi设备的类型、跟踪分析不同设备类型的网络流量模型，这样才能更好的优化网络配置，提供更有针对性的网络增值服务[3-5]。目前对WiFi信号的监测方法主要有两种：一种是利用专用的探测设备主动监测，另外一种是非侵入的被动监测。主动监测方面，一些研究利用beacon节点进行室内行人或WiFi设备的定位跟踪，但部署beacon节点成本较高[6-7]。于是人们转向了对WiFi信号被动监测的研究。

文献[8]通过WiFi被动监测，融合WiFi信号、手机传感器以及室内拓扑进行室内行人的导航。文献[9]和文献[3]分别利用了WiFi通信时的探测请求帧和信道状态信息来评估人群拥挤程度。文献[10]通过采集包含在Probe帧里的SSID(Service Set Identifier)列表信息构建指纹库，以此判断两个设备的用户是否存在社会关系。文献[11]和文献[12]都采用了相似的方法分析基于用户行为的时空相似度，进而推测用户之间的社交关系。文献[13]构建了一个可以在超市卖场实现对用户按游览区域分类的系统GruMon，该系统主要利用了手机自带传感器的数据，如重力加速度传感器、指南针、气压计等，这篇文献做的工作与本文的设备分类有相似之处，但本文主要关注的是WiFi被动监测的数据，这种分类方式无需从用户手机上获取相关的传感器信息，是一种非入侵、低成本的方法。文献[14]通过长时间被动监测手机WiFi通信中的参数来

* 基金项目：国家自然科学基金项目(61379127)资助
Supported by the National Natural Science Foundation of China (61379127)
收稿日期：2018-04-16；修订日期：2018-11-05
作者简介：姜明星(1984-)，男，讲师。E-mail：jiangmx@ouc.edu.cn

发现用户发生的地点相关事件，进而推测他们的社交关系，文中采集数据的方式及处理的数据与本文中比较相似，但本文更关注设备的分类及设备之间的关系。本文提出了一种基于非入侵 WiFi 信号监测的设备自动分类机制，该机制可以快速准确的识别设备类型。

通过普通的 WiFi 探测工具可以不间断的获取设备无线通信中的相关信息，比如接收信号强度 RSS、发送设备(源)MAC 地址、时间戳等。我们把这些可以长期获取并且稳定可靠的通讯参数记录下来，分析特定设备在一定时间范围内的接收信号强度变化趋势和规律，提取每类设备的特征，用于构建相应的特征向量，再使用监督学习的方法训练不同的分类器，从而实现诸多不同类型设备的分类。图 1 是一个手机、一台笔记本电脑和一台无线路由器一天接收信号强度的变化图，不难看出这三种设备每天的接收信号强度变化趋势是不一样的并且有一些各自独有的特征。

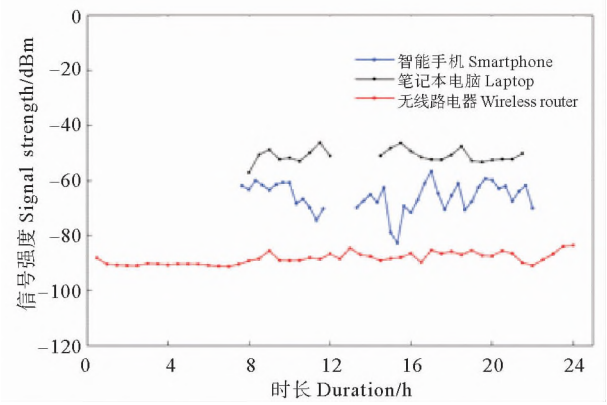

图 1　三种典型 WiFi 设备每日接收信号强度变化图

Fig.1　Daily RSS variations of 3 typical WiFi devices

1　数据采集预处理

由于 WiFi 探测设备只能获取附近一定范围内的无线设备信号，所以监测范围内的无线设备数量，在不同时间段内也不尽相同。我们关注的是那些持续且稳定的无线信号，即那些经常出现在监测范围内的设备，为此我们专门设计一个过滤器来筛选出这些“常客”。无线设备筛选工作是基于对设备长期监测数据的分析，图 2 显示了访客和常客一周内的在线情况(即有信号的时段)，访客对应的在线时间明显少于常客。实际筛选时，我们统计每个设备接收信号强度的平均值、设备累计在线天数和设备每天在线时长，具体统计规则为：

(1)接收信号强度均值：统计某设备在所有监测时间内接收信号强度的平均值。

(2)在线天数：如果一天之内监测到来自某设备的信号，则该设备当天在线。

(3)在线时长：如果在连续的 5 min 之内监测到了某设备的信号，则认为该设备这 5 min 在线，以此 5 min 为单位，统计该设备一天的累计在线时长。

根据经验，我们会筛除那些统计周期内累计在线天数小于 5 天和每天在线时长小于 5 min 的设备，因为这些数据基本来自于那些不经常出现在监测范围内的设备，设备对应的用户也很可能是一些访客。此外，我们还会筛除那些接收信号强度平均值小于－110 dB 的设备，因为这些设备即使能长期获取监测数据，但是由于距离监测设备太远，接收信号强度变化规律不明显，很难提取有效特征。

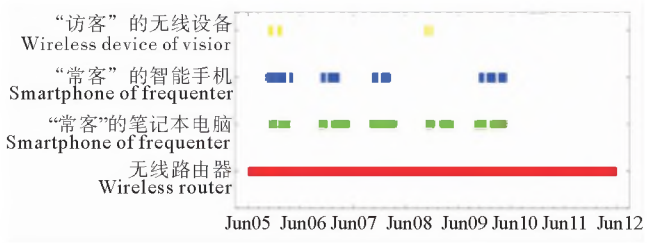

图 2　一周时间内各类设备的在线情况图

Fig.2　Active span for different devices during a week

2　设备分类算法

尽管部分无线 WiFi 设备的类型，可以通过调用网上一些基于 MAC 地址的查询服务来获取，但实际上大多数的设备类型信息并不能据此方法准确得出，究其原因大致有两种情况：(1)部分设备制造商没有对他们不同类型的产品采用特定的 MAC 地址分类策略；(2)另外一些设备制造商出于商业保密的考虑故意混淆或隐藏他们的 MAC 地址分类策略。所以要实现准确的无线设备类型分类，完全依靠网上现有的查询服务无法实现。

本节主要通过分析长期监测的无线设备的信号数据，提取不同类型设备的特征，使用机器学习的方法，完成设备分类。这种方法的好处是不需要主动读取无线设备的数据即可完成高精度的分类。这里我们根据 WiFi 无线设备的特征，把它们分成 3 类：智能手机类 Mobile Phone，简称 MP 类；笔记本或平板电脑 Laptop，简称 LT 类；和静止的 WiFi 设备 Stationary Machine，如无线路由器或使用 WiFi 联网的台式机，简称 SM 类。

2.1 事件识别步骤

本节提到的三种 WiFi 设备，来自相同类型设备的数据包在很多方面是相似的。基于此，我们分析提取每种类型设备的特征用于对设备分类，主要特征如下：

2.1.1 在线率　　统计每个 MAC 地址对应的设备每天

的在线率 α。在线率即每天累计在线时间除以 24 h，如公式 1 所示。我们截取了两周时间的监测数据计算三种设备的在线率。如图 3 所示，可以看出，不同类型设备的在线率存在明显差异，对于静止的无线设备（SM 类），它们的在线率一般接近 100%，而其他两类设备在线率一般在 40%～60%之间。在线时间 T_o 的统计方法为：以 5 min 为单位，监测是否收到该设备的数据包，如果收到则记该设备这 5 min 在线，否则记为离线。以此统计设备一天（24 h，即 288 个单位时间）的在线时间。如公式 2 所示，p_i 代表第 i 个时间段内，设备是否在线，其值为 0 或 1。

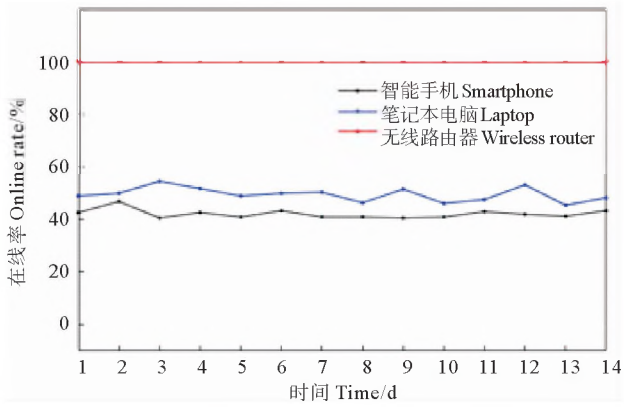

图 3 两周时间内各类设备的在线率情况

Fig.3 Daily online rate of different devices during 2 weeks

$$\alpha = T_o/288。 \tag{1}$$

$$T_o = \sum_{i=1}^{288} 5 \cdot p_i 。 \tag{2}$$

2.1.2 接收信号强度（RSS） 接收信号强度反映了探测设备接收到数据包的能量，主要和发送方与接收方的距离有关，还有一些其他因素，诸如是否存在障碍物、天线角度等。可以预见地，可移动无线设备接收信号强度在时域的波动率一般要高于笔记本电脑类的，而这两类无线设备的波动率要高于那些静止的 WiFi 设备。我们引入接收信号强度的标准差 σ_r 和极差 R 作为分类特征，具体计算方法有如下公式：

$$\sigma_r = \sqrt{\frac{\sum_{i=1}^{N}(r_i - \bar{r})^2}{N}} 。 \tag{3}$$

$$R = r_{\max} - r_{\min}。 \tag{4}$$

其中：$\bar{r}$ 是指设备一天接收信号强度的均值；N 为一天内接收该设备数据包的总数。各类设备标准差的具体差异如图 4 所示：几乎所有的静止 WiFi 设备的标准差小于 5 dBm，有 80%左右的笔记本电脑的标准差小于 5 dBm，相对而言标准差小于 5 dBm 的移动手机只有不到 50%。

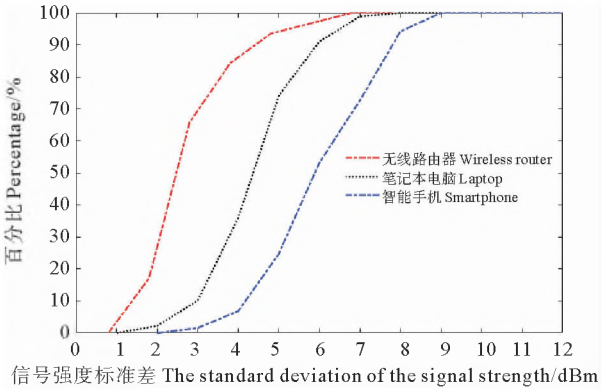

图 4 三类设备接收信号强度标准差 CDF

Fig.4 CDF of the standard deviation of the RSS for different devices

2.1.3 数据包间隔 不同设备发送数据包的频率是不一样的，SM 类设备由于有持续稳定的电源供应，数据包发送频率最高而且发送周期也非常整齐相近；LT 类设备在工作时数据包发送频率较高，发送周期也比较相近；MP 类设备由于只能依靠自带电池供电且处于节能考虑都自带休眠机制，这导致手机待机状态下的数据包发送数目明显少于工作状态（手机解锁，屏幕常亮的状态）。为了描述这些差异，我们引入数据包平均发送周期（$\overline{T}$）和发送周期的标准差（σ_T）作为另外的分类特征，具体定义见如下公式：

$$\overline{T} = \frac{1}{N-1}\sum_{i=1}^{N-1} p_i 。 \tag{5}$$

$$\sigma_T = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(p_i - \overline{T})^2} 。 \tag{6}$$

这里记来自同一设备的数据包的接收时间为序列 $T=\{t_1, t_2, \cdots, t_N\}$，则数据包间隔 $p_i = t_{i+1} - t_i$，考虑到一天内手机和笔记本电脑的数据包或有间断，当实际计算发送周期的均值和标准差时，过滤掉相邻时间点间隔大于 5 min 的数据点。

2.2 层次化分类

由于提取的三类设备的特征并非各自独立，为了达到更好的分类效果，本文采用层次化分类方法，具体过程为：首先设计静态设备（SM、LT 类）和移动设备（MP 类）分类器，用于识别数据样本来源于静态设备（无线路由设备）还是移动设备（手机、笔记本电脑）；然后在第一次分类的基础上对移动设备进行二次分类，设计用于识别数据样本是手机还是笔记本电脑的分类器。针对两次分类，分别选择不同的特征构建特征向量。实验证明这种层次化分类方法比直接对无线设备进行三分类的平均准确率要高。

3　实验及讨论

3.1 实验设置

本实验通过 WiFi 探针进行数据采集，采集地点为普通房间，房间内有若干办公台位。以天为单位，进行数据采集并存储。采集持续 32 d，记录采集时间、MAC 地址以及接收信号强度，本次采集原始数据共计 5 218 条。进行分析前，首先对数据进行了过滤，过滤标准如下：保留以天为单位的完整数据（全天只有一类设备的数据、全天没有数据、全天只有半天的数据，以上这三种类型的数据去除），过滤后剩余数据 666 条。经过数据预处理阶段，得到最终的实验样本共计 369 条。

样本标签设置：第一次分类：0 表示无线路由器，1 表示手机和笔记本；第二次分类：2 表示笔记本，3 表示手机。

训练平台：采用 Matlab R2016a 以及其软件自带分类工具箱，参数默认，并采用十折交叉验证方式。

3.2 实验过程

3.2.1 *层次化分类*　每次选用不同特征集对标签数据进行分类，最后得出层次化分类的平均准确率。训练算法：Fine Gaussian SVM（FG-SVM），Cubic KNN（C-KNN），Simple Tree（S-Tree），Linear Discriminant（LD）。

第一次分类实验：

（1）本次实验主要对过滤后样本集进行训练，从而实现静态设备和移动设备的分类。实验使用了 5 个特征，分别为时间在线率（Online Time Rate，OTR），信息发送时间间隔的平均值（Average Sending Time Interval，A-STI），信息发送时间间隔的标准差（Sending Time Interval Standard Deviation，STI-SD），接收信号强度的标准差（RSS Standard Deviation，RSS-SD），接收信号强度的极差（RSS Range，RSSR）。

（2）将得到的全部设备的上述特征以及标签文件导入到 Matlab R2016a 中，进行分类。

（3）在分类的过程中，首先进行特征验证，单个特征对应的分类准确率如表 1 所示。

表 1　单个特征验证结果

Table 1　Verification results of single feature

特征 Feature	算法 Algorithm			
	FG-SVM	C-KNN	S-Tree	LD
OTR	97.4	96.8	96.8	97.1
A-STI	90.4	89.8	89.9	77.5
STI-SD	89.2	90.5	88.6	76.7
RSS-SD	89.2	80.0	79.4	72.5
RSSR	75.8	70.9	75.2	72.2

从表中可以看出，分类用到的五个特征准确率均达到 70%以上，可见所选特征都是有效特征。接下来使用这五个特征构建特征向量进行分类，结果如表 2 所示。

表 2　五个特征验证结果

Table 2　Verification results of all features

算法 Algorithm	FG-SVM	C-KNN	S-Tree	LD
准确率 Accuracy	98.6	98.9	98.3	97.3

第二次分类实验：

（4）针对第一次分类中被归类为移动设备的数据集进行再训练，以实现手机和笔记本电脑的分类。这次实验使用了另外 5 个特征，分别为信息发送时间间隔的平均值（Average Sending Time Interval，A-STI）、信息发送时间间隔的标准差（Sending Time Interval Standard Deviation，STI-SD），时间间隔标准差/时间间隔平均值（STI-SD/A-STI），接收信号强度的标准差（RSS Standard Deviation，RSS-SD），接收信号强度的极差（RSS Range，RSSR）。

（5）将得到的全部设备的上述特征以及标签文件导入到 Matlab R2016a 中，进行分类。

（6）在分类的过程中，首先进行特征验证，单个特征验证对应的分类准确率如表 3 所示。

表 3　单个特征验证结果（第二次分类）

Table 3　Verification results of single feature in the second classification

特征 Feature	算法 Algorithm			
	FG-SVM	C-KNN	S-Tree	LD
A-STI	81.9	79.7	81.0	75.1
STI-SD	81.4	82.1	81.8	74.8
STI-SD/A-STI	81.1	81.6	82.9	74.5
RSS-SD	81.4	82.4	81.6	75.6
RSSR	76.8	76.4	77.5	77.5

从表中可以看出，分类用到的五个特征准确率均达到 70%以上，所选特征即为有效特征。接下来使用这个 5 个特征构建特征向量进行二次分类，结果如表 4 所示。

由于样本数据不均衡，因此本工作采用平均准确率作为本文分类工作的评价标准，在两次实验的基础上，求出平均准确率，具体的方法如下：

表4 五个特征验证结果(第二次分类)

Table 4 Verification results of all featuresin (the second classification)

算法 Algorithm	FG-SVM	C-KNN	S-Tree	LD
准确率 Accuracy	87.5	89.2	87.5	89.2

(1) 第二次实验手机分类准确率=提取出手机的准确条数/提取出手机的总条数。

(2) 第二次实验笔记本电脑分类准确率=提取出笔记本电脑的准确条数/提取出笔记本电脑的总条数。

(3)第一次实验移动设备分类准确率=提取出移动设备的准确条数/提取出移动设备的总条数。

(4)第一次实验无线路由器分类准确率=提取出无线路由器的准确条数/提取出无线路由器的总条数。

(5)平均准确率=((第二次实验手机分类准确率+第二次实验笔记本电脑分类准确率)*第一次实验移动设备分类准确率+第一次实验无线路由器分类准确率)/3。

在此过程中,分别使用SVM、KNN、Decision Tree (DT)、Discriminant Analysis(DA)等算法进行了平均准确率的计算,其结果如表5所示。

表5 层次化分类平均准确率

Table 5 Average accuracy of hierarchical classification

算法 Algorithm	第一次分类 First	第二次分类 Second			准确率 Accuracy
	移动设备 Mobile device	无线路由器 WiFi	手机 Cell phone	笔记本 Laptop	
SVM	99.19	98.99	95.31	81.52	91.46
DT	97.84	97.31	93.86	66.3	84.67
KNN	98.66	98.32	92.42	78.26	88.90
DA	98.91	94.94	92.77	78.26	88.04

3.2.2 *层次化分类有效性验证* 为了验证本文提出的层次化分类策略,现将层次化分类中两次分类用到特征的并集作为三分类算法的特征集,使用和层次化分类相同的数据集合,逐一训练分类得出每种三分类算法的平均准确率,这里平均准确率等于三种设备准确率的平均值(见表6)。

表6 三分类及层次化分类平均准确率

Table 6 Average accuracy of 3-class classification and hierarchical classification

算法 Algorithm	SVM	KNN	DT	DA
三分类 3-Class	80.41	85.08	82.84	82.05
层次化分类 Hierarchical	91.46	88.90	84.67	88.04

从上表可见使用本文提出的层次化分类机制,不论采用哪种分类算法,都能得到更高的平均准确率。

4 结语

本文通过对无线通信技术中最常见的WiFi接收信号强度的长期非入侵式监测,探究具体无线设备的物理属性,应用机器学习的成熟算法,实现了无线设备的识别及分类。本文提出的设备类型识别机制可以应用在无线网络负载分析、网络部署优化等方面,有助于部署更具个性化的无线网络增值服务。本文的工作为今后无线网络的管理优化提供了必要条件,为探索物联网设备间的关联性做出了有益尝试。

参考文献:

[1] CISCO I. Cisco visual networking index: forecast and methodology 2014-2019, white paper[EB/OL]. [2015-06-18][2018-05-29] http: //www. datacenterconsulting. com/2015/06/18/cisco-visual-networking-index-forecast-and-methodology-2014-2019.

[2] U Kumar, J Kim, A Helmy. Comparing Wireless Network Usage: Laptop vs Smart-phones[C]. Miami: The 19th Annual International Conference on Mobile Computing & Networking, 2013: 243-246.

[3] Papapanagiotou I, Nahum E M, Pappas V. Smartphones vs. laptops: comparing web browsing behavior and the implications for caching[J]. ACM Sigmetrics Performance Evaluation Review, 2012, 40: 423-424.

[4] Wei N, Valler, H V Madhyastha, et al. A Behavior-Aware Profiling of Handheld Devices[C]. HongKong: In Computer Communications (INFOCOM), 2015: 846-854.

[5] M Zhu, Z Zeng, L Wang, et al. A Measurement Study of a Campus Wi-Fi Network with mixed Handheld and Non-handheld Traffic[C]. Halifax: Electrical and Computer Engineering (CCECE), 2015: 848-853.

[6] Sichitiu M L, Ramadurai V. Localization of Wireless Sensor Networks with a Mobile Beacon[C]. Dalian: IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, 2013: 174-183.

[7] Acuna V, Kumbhar A, Vattapparamban E, et al. Localization of Wi-Fi Devices Using Probe Requests Captured at Unmanned Aerial Vehicles[C]. San Francisco: Wireless Communications and Networking Conference, 2017: 1-6.

[8] Wang X, Jiang M, Guo Z, et al. An indoor positioning method for

smartphones using landmarks and PDR[J]. Sensors, 2016, 16 (12): 2135.

[9] Luzio A Di, Mei A, Stefa J. Mind your probes: De-Anonymization of Large Crowds Through Smartphone Wi-Fi Probe Requests[C]. San Francisco: Computer Communications (INFOCOM), 2016.

[10] Rouveyrol P, Raveneau P, Cunche M. Large Scale Wi-Fi tracking using a Botnet of Wireless Routers[C]. Philadelphia: Workshop on Surveillance & Technology, 2015.

[11] Fukuzaki Y, Mochizuki M, Murao K, et al. Statistical Analysis of Actual Number of Pedestrians for Wi-Fi Packet-based Pedestrian Flow Sensing[C]. New York: the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2015 ACM International Symposium on Wearable Computers, 2015: 1519-1526.

[12] Chon Y, Kim S, Lee S, et al. SensingWi-Fi Packets in the Air: Practicality and Implications in Urban Mobility Monitoring[C]. New York: the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2014: 189-200.

[13] Maier G, Schneider F, Feldmann A. A first look at mobile handheld device traffic[J]. Passive and Active Measurement, Springer, DOI: 10.1007/978-3-642-12334-4-17.

[14] Hong H, Luo C, SocialProbe M C: Understanding Social Interaction Through Passive WiFi Monitoring[C]. Hiroshima: Computing, Networking and Services, ACM, 2016: 94-103.

Identification Mechanism of Device Type Based on Wi-Fi Signal Strength Monitoring

JIANG Ming-Xing[1], WANG Xi[2], GUO Zhong-Wen[2], WANG Jin-Xin[2]

(1.Teaching Center of Fundamental Courses, Ocean University of China, Qingdao 266100, China; 2. College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China)

Abstract: With the development of wireless communication technologies, Wi-Fi devices are becoming widely popular and grow rapidly in the amount andvariety and as long as the Wi-Fi functionality of such device is switched on, signals containing lots of information can be monitored. The network traffic problem due to numerousWi-Fi devices has been increasinglystudied by researchers. However tracking network traffic must be built on prioracquaintance of device type. At present, there are not so many studies on recognition of Wi-Fi devices and they all rely on beacon nodes to actively collect wireless signals. In this paper, based on non-invasive collecting of Wi-Fi signal data, a method for classifying Wi-Fi enabled devices is proposed. By constantly monitoring some key parameters in Wi-Fi communication, such as RSS (Received Signal Strength), MAC address, timestamp, etc. we collect a lot of long-term data and divide them in units of 24 hours a day. According to the behavioral patterns of such devices' carriers, we extract different types of devices' features, and represent each device with a feature vector. Then different classifier algorithms are adopted to categorize these common Wi-Fi enabled devices, including smartphones, laptops and wireless routers. According to result of the experiment, the type recognition mechanism proposed in this paper achieves better performance, applying todifferent classification algorithms for hierarchical classification.

Key words: Wi-Fi monitoring; Received Signal Strength (RSS); machine learning; hierarchical classification

责任编辑　徐　环