

不依赖位置坐标的室内 Wi-Fi 网络覆盖度量方法

谢泽锋¹, 陈伟栋¹, 黄黎霞^{2,3}, 顾一帆^{2,3}, 张博钧⁴, 全智^{2,3}

(1. 中国移动通信集团广东有限公司, 广东 广州 510623; 2. 深圳大学射频异质异构集成全国重点实验室, 广东 深圳 518060;
3. 深圳大学电子与信息工程学院, 广东 深圳 518060;
4. 中国信息通信研究院南方分院 (深圳信息通信研究院), 广东 深圳 518035)

摘要: 3GPP 在版本 16 (R16, Release 16) 中升级了最小化路测 (MDT, minimization of drive test) 技术, 提出移动终端可利用 4G/5G 网络自主上报 Wi-Fi 信号的接收信号强度指示 (RSSI, received signal strength indicator), 为运营商度量 Wi-Fi 网络的覆盖率带来了可能性。然而, 现有基于 MDT 技术的网络覆盖度量方法严重依赖 GPS 提供的位置坐标, 但全球定位系统 (GPS, global positioning system) 不能提供室内精准定位, 无法用于室内 Wi-Fi 网络的覆盖度量。为此, 提出了一种不依赖位置坐标的 RSSI 聚类方法, 充分利用室内相近位置 RSSI 的统计相似性, 区分不同位置的 RSSI 测量差异, 在无位置坐标条件下准确估计出室内 Wi-Fi 网络的覆盖率。实验结果表明, 所提方法估计的覆盖率与基于真实位置坐标测量的覆盖率相近, 度量准确度明显优于现有的其他方法。

关键词: 网络覆盖率; Wi-Fi 网络; 最小化路测; 聚类算法; 接收信号强度指示

中图分类号: TN915.857

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2024.00397

Indoor Wi-Fi coverage measurement without spatial coordinates

XIE Zefeng¹, CHEN Weidong¹, HUANG Lixia^{2,3}, GU Yifan^{2,3}, ZHANG Bojun⁴, QUAN Zhi^{2,3}

1. China Mobile Group Guangdong Co., Ltd., Guangzhou 510623, China

2. State Key Laboratory of Radio Frequency Heterogeneous Integration, Shenzhen University, Shenzhen 518060, China

3. College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China

4. South Institute of China Academy of Information and Communications Technology (Shenzhen Academy of Information and Communications Technology), Shenzhen 518035, China

Abstract: In R16, 3GPP proposed that user devices could utilize the 4G/5G cellular networks to report the received signal strength indicator (RSSI) of Wi-Fi signal based on the existing minimization of drive test (MDT) technology, making it possible to measure the coverage probability of indoor Wi-Fi networks. However, existing measurement methods of network coverage probability based on the MDT require the spatial coordinates provided by the GPS. As GPS has poor indoor localization accuracy, it is not able to be applied to indoor Wi-Fi networks. A measurement method for network coverage probability based on clustering without spatial coordinates was proposed. The proposed method could distinguish RSSI measurement reported on different locations, with the fact that their statistics were similar at similar locations. The coverage probability was accurately measured by utilizing the clustering results without knowing the spatial coordinates. Experimental results show that the coverage probability measured by the proposed method is very close to the probability measured by the known spatial coordinates, and the accuracy is much higher than existing methods.

Key words: coverage probability, Wi-Fi network, minimization of drive test, clustering algorithm, received signal strength indicator

收稿日期: 2024-03-26; 修回日期: 2024-06-15

通信作者: 张博钧, zhangbojun@caict.ac.cn

基金项目: 国家自然科学基金项目 (No. 62201361)

Foundation Item: The National Natural Science Foundation of China (No. 62201361)

0 引言

随着无线通信网络的发展, 第三代合作伙伴计划(3GPP, 3rd Generation Partnership Project)主导的4G/5G移动蜂窝网络通信技术与IEEE主导的Wi-Fi网络通信技术正在协同并进、互补融合^[1-3]。4G/5G通信技术主要面向室外区域的覆盖, 而Wi-Fi技术主要应用于室内区域的覆盖^[4-5]。围绕4G/5G蜂窝网络和Wi-Fi网络, 移动手机业务和家庭宽带业务成为运营商的两大核心业务场景。为此, 运营商不仅需保障4G/5G网络的覆盖质量, 还需关注Wi-Fi网络的覆盖质量。网络覆盖的好坏可由覆盖率(coverage probability)评判, 其定义为网络空间中任意位置接收信号质量大于弱覆盖门限的概率^[6-8]。准确度量网络覆盖率是进一步诊断网络问题、优化网络参数等的根本前提。

早期4G/5G网络的覆盖率度量主要基于路测(drive test)^[9-10]。在该方法中, 测试人员驾驶车辆在网络各处采集与接收信号质量相关的参数, 以形成测试报告并计算覆盖率, 报告内容包括测试位置、信干噪比(SINR, signal to interference plus noise ratio)、接收信号强度指示(RSSI, received signal strength indicator)等^[11]。然而, 这种依赖于人工测量的方式存在诸多问题, 如测量结果的客观性不足、测量效率低、数据采集量有限等。为应对这些问题, 3GPP在版本10(R10, Release 10)中提出了最小化路测(MDT, minimization of drive test)技术^[12]。这项技术通过用户终端向基站自主上报测量位置及数据的方式, 实现了大规模数据收集, 引起了学术界和产业界的广泛关注。针对MDT技术的研究包括网络覆盖率的高效度量^[13-16]、网络问题诊断^[17-19]、网络参数及覆盖率优化^[20-23]等。

运营商虽可基于MDT技术有效地获取室外4G/5G网络的覆盖率, 但仍然无法获取用户室内Wi-Fi网络的覆盖率。室内用户的通信体验很大程度上取决于这“最后一段”Wi-Fi网络, 因此室内Wi-Fi网络的覆盖情况至关重要。2021年, 3GPP在提出的R16中指出, 室内Wi-Fi设备可以利用4G/5G网络的广泛覆盖, 基于现有的MDT技术进一步上报Wi-Fi信号的RSSI^[24]。这一修订为运营商度量室内Wi-Fi网络的覆盖率提供了可行性。

然而, 目前基于MDT的室外4G/5G网络覆盖

率评估方法依赖于上报数据在空间中的位置坐标以准确计算网络覆盖率^[13-23], 无法被用于室内Wi-Fi网络中。在室外4G/5G网络中, 该位置坐标由全球定位系统(GPS, global positioning system)提供的经纬度表示。由于GPS定位在室内空间中存在较大误差^[25-28], 很难区分具体楼层及房间, 无法被用于室内Wi-Fi网络中。此外, 现有Wi-Fi协议也不支持终端位置的度量, 无法提供准确的室内位置坐标^[29]。因此, 如何在无位置坐标的条件下, 仅通过上报的RSSI数据准确度量室内Wi-Fi网络的覆盖率成为本研究所面临的核心挑战。

针对上述挑战, 提出了一种基于聚类(clustering)的室内Wi-Fi网络覆盖率度量方法, 该方法充分利用室内相近位置RSSI的统计相似性, 将室内覆盖率求解的问题重构为一个针对上报RSSI数据的聚类问题(clustering problem)。通过这一方法, 可以在无位置坐标条件下准确地估计出室内Wi-Fi网络的覆盖率。实验结果表明, 所提方法在无位置坐标条件下估计的覆盖率与已知位置坐标条件下的真实覆盖率非常接近, 且其覆盖率度量准确度明显高于其他估计方法。

1 系统框架与问题描述

1.1 系统框架

MDT基站回传终端Wi-Fi测量数据的网络覆盖测试系统如图1所示, 本文考虑一个MDT基站回传终端Wi-Fi测量数据的网络覆盖测试系统。在该系统中, 终端用户根据3GPP在R16中提出的MDT技术^[24], 自主采集室内Wi-Fi网络的RSSI。所采集的RSSI样本不包含室内位置坐标, 且采集路径及采集位置均不可控。终端用户将多个RSSI数据封装成MDT测量报告, 周期性地通过4G/5G网络经由基站回传至云服务器。其中, 第 p 个MDT测量报告回传的RSSI数据可表示为

$$\mathbf{r}_p = (r_{p,1}, r_{p,2}, \dots, r_{p,Q}) \quad (1)$$

其中, Q 为每个MDT测量报告中RSSI数据的个数, 可由基站预先指定^[30], $r_{p,q}$ 为第 p 个MDT测量报告中的第 q 个RSSI样本。

为了精确度量室内Wi-Fi网络的覆盖率, 云服务器可利用多个存储的MDT测量报告, 若云服务器的存储能力为 P , 则所有可被用于覆盖率度量的RSSI数据可表示为

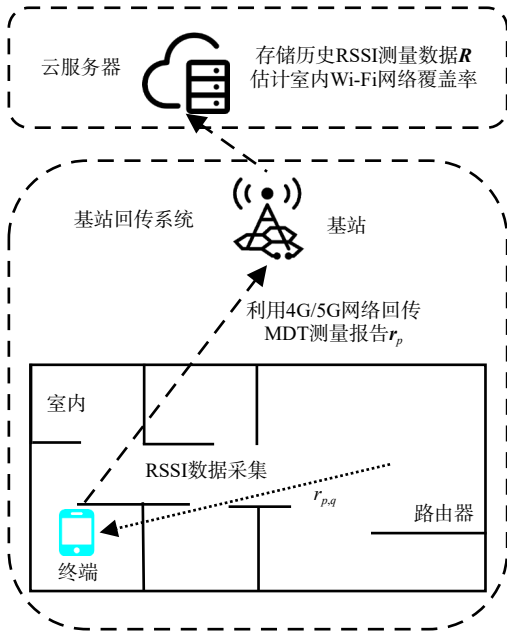


图1 MDT基站回传终端Wi-Fi测量数据的网络覆盖测试系统

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_P \end{pmatrix} = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,Q} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ r_{P,1} & r_{P,2} & \cdots & r_{P,Q} \end{pmatrix} \quad (2)$$

其中，矩阵 \mathbf{R} 的维度为 $P \times Q$ ，每一行代表一个MDT测量报告，每一列则代表该MDT测量报告中每个RSSI的样值。接下来，云服务器将根据近期所有上报的RSSI数据，即矩阵 \mathbf{R} ，度量室内Wi-Fi网络的覆盖率。

1.2 问题描述

网络覆盖率是衡量网络整体性能的重要指标之一，其定义为接收信号质量大于弱覆盖门限的概率^[6-8]。由于3GPP提出的MDT技术只支持室内Wi-Fi信号的RSSI数据上报，本文使用RSSI表示接收信号质量。基于此，室内Wi-Fi网络的整体覆盖率可被表示为

$$\eta = \mathbb{P}(r > \delta) = \int_{\delta}^{\infty} p_r(x) dx \quad (3)$$

其中， r 为室内空间中任意位置的RSSI，为一个随机变量； δ 为基于RSSI指标的弱覆盖门限； $p_r(x)$ 表示随机变量 r 在空间中的概率密度函数（PDF, probability density function）。

由于终端用户RSSI采集位置的不确定性，概率密度函数 $p_r(x)$ 很难通过所有测量数据即 \mathbf{R} 估计，式(3)的积分无法直接求解。为了解决这一问题，现有测量方法通过栅格化的方式计算。将整个室内

空间细化为 K 个栅格，当栅格数目足够大时，即每个栅格的面积足够小时，式(3)可被表示为

$$\eta = \frac{1}{K} \sum_{k=1}^K \eta_k = \frac{1}{K} \sum_{k=1}^K \frac{m_k}{n_k} \quad (4)$$

其中， η_k 表示栅格 k 的覆盖率， n_k 表示该栅格中采集的RSSI样本总数， m_k 表示该栅格中RSSI大于弱覆盖门限 δ 的样本个数，它们与 \mathbf{R} 中每个RSSI样本所属的栅格相关。

通过式(4)不难发现，为了得到 n_k 和 m_k 以计算覆盖率，需要上报RSSI数据的位置坐标以确定其所属栅格。利用GPS提供的经纬度位置坐标，式(4)的方法可被用于室外4G/5G网络的覆盖度量。但室内Wi-Fi终端上报的RSSI数据不包含位置坐标，无法通过式(4)准确地计算覆盖率。在无位置坐标条件下，将整个室内空间视为一个大栅格进行估计为一个粗略的覆盖度量方法。

$$\tilde{\eta} \approx \frac{M}{PQ} = \frac{\sum_{k=1}^K m_k}{\sum_{k=1}^K n_k} \quad (5)$$

其中， PQ 表示服务器中存储的所有RSSI样本个数，即矩阵 \mathbf{R} 中的元素个数。 M 表示所有RSSI大于弱覆盖门限 δ 的样本个数。

由于式(5)的估计方法没有考虑栅格划分和RSSI样本位置坐标的影响，将产生极大度量误差。本文举例更好地阐述这一点，RSSI测量结果样例如图2所示。在该例子中，整个室内空间被划分为4个栅格，每个栅格给出了采集的RSSI样本总数 n 及RSSI大于弱覆盖门限的样本个数 m 。将该测量结果代入式(4)和式(5)分别求得覆盖率为75%和95%。不难看出，式(4)的结果75%反映了该室内Wi-Fi网络的真实覆盖率，而式(5)的结果95%则存在明显偏差。因此，亟须研究无位置坐标条件下的室内Wi-Fi网络覆盖率度量新方法，使得估计覆盖率 $\tilde{\eta}$ 趋近于式(4)求得真实覆盖率。该问题的难点在于式(4)中的真实覆盖率依赖于RSSI样本的位置坐标，而本文提出的覆盖率度量方法只基于无位置坐标下的所有上报数据 \mathbf{R} 。

2 基于聚类的室内Wi-Fi网络覆盖率度量方法

本节首先将无位置坐标下的覆盖率求解问题重构为一个聚类问题，再设计相应的聚类算法并根据

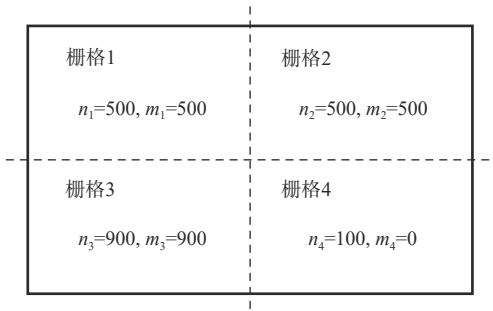


图2 RSSI测量结果样例

聚类后的结果估计室内Wi-Fi网络的覆盖率。

2.1 问题重构

MDT协议测量报告的上报周期间隔通常为毫秒级别^[30], 终端Wi-Fi设备不会在单个上报周期内发生显著位移, 故其每次上报的RSSI测量报告 r_p 中的 Q 个样本可视为同一栅格所采集。由于RSSI在室内不同位置的统计特性存在差异^[29], 不同栅格采集的报告 r_p 的统计特性也不同, 可以利用这一点区分不同栅格中上报的 r_p 进而准确估计室内Wi-Fi网络的覆盖率。这样的区分方式并不依赖 r_p 的具体位置坐标, 只关注其本身的统计特性, 因此适用于无位置坐标的前提假设。

聚类问题重构示意图如图3所示, 本文将无位置坐标条件下的室内Wi-Fi网络覆盖率度量问题重构为一个针对服务器中所有 P 个MDT测量报告 r_p 的聚类问题。在图3的例子中, Wi-Fi终端在某一栅格 k 中共测得两个测量报告 r_i 和 r_j 。根据 r_i 和 r_j 中 Q 个RSSI样本统计特性的相似性, 可将两个测量

报告通过聚类算法聚为某一类, 用聚类 l 表示, 该聚类算法的设计将在下文阐述。因此, 栅格 k 与聚类 l 中包含的RSSI样本完全相同, 它们各自所对应的覆盖率也相同。以此类推, 当每个栅格中所采集的 r_p 均能被聚为一类时, 则每个聚类所对应的覆盖率等同于式(4)中每个栅格所对应的覆盖率。因此, 无位置坐标条件下, 基于聚类方法的室内Wi-Fi网络覆盖率可被表述为

$$\tilde{\eta} \approx \frac{1}{L} \sum_{l=1}^L \tilde{\eta}_l = \frac{1}{L} \sum_{l=1}^L \frac{\tilde{m}_l}{\tilde{n}_l} \quad (6)$$

其中, L 表示聚类个数, 理想情况下等于式(4)中的栅格个数 K 。 \tilde{n}_l 和 \tilde{m}_l 分别表示聚类 l 中RSSI样本的个数与大于弱覆盖门限样本的个数, 理想情况下与式(4)中某一栅格的 n_k, m_k 对应。

显然, 式(6)的度量准确度依赖于聚类算法的聚类准确度。值得强调的是, 聚类算法在对MDT测量报告 r_p 进行聚类时, 无须决定 r_p 的绝对位置坐标, 只需要基于 r_p 的统计特性判断其与其他测量报告之间的相对关系, 因而通常能实现较高的聚类精度。此外, 一定的聚类偏差并不会影响式(6)的准确度, 以图2例, 当栅格1的部分上报数据被聚类到栅格2的上报数据中时, 式(6)的覆盖率仍然等同于真实值75%, 这一点将在后续的实验中进行验证。相反, 若不采纳聚类方法而根据式(5)粗略地估计覆盖率, 所得的结果95%存在明显偏差。接下来, 本文将针对服务器中所有 P 个MDT测量报告 r_p 的聚类问题, 设计相应的聚类算法, 从而可以采

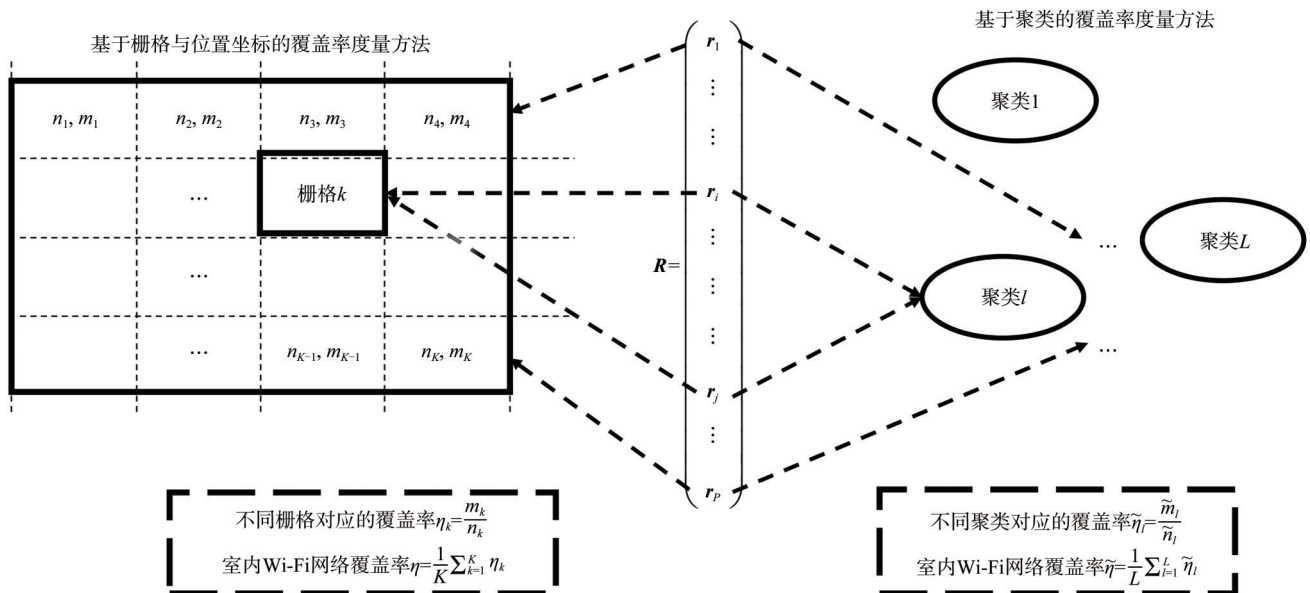


图3 聚类问题重构示意图

用式(6)准确地估计覆盖率。

2.2 聚类算法

聚类算法将根据 r_p 的统计特性进行聚类以区分不同栅格中上报的 r_p ，因此，首先需要定义 r_p 的统计特征。为了充分挖掘上报数据 r_p 的统计特征，由于每个 r_p 中包含 Q 个 RSSI 样值，本文将 r_p 中 Q 个样值的均值以及中值作为特征，分别表示为 x_p 和 y_p 。基于这两个特征，接下来针对所有 P 个 MDT 测量报告的二维特征向量 $z_p = (x_p, y_p)$ 进行聚类，并将每个特征向量视为二维特征空间中的一个点。

聚类算法在学术界已被广泛研究，其中， K 均值 (K -means)、DBSCAN (density-based spatial clustering of applications with noise) 和 Mean Shift 是 3 个经典的聚类算法^[31-33]，值得借鉴。 K -means 算法需要已知聚类个数，即栅格数量 K 。由于不同规模的室内 Wi-Fi 网络应当采纳不同的栅格数量以准确地度量覆盖率，且服务器在估计覆盖率时未知该栅格数量，因此 K -means 算法并不适用于本文所研究的问题。与 K -means 算法不同，DBSCAN 和 Mean Shift 算法利用数据密度进行聚类，因而无须已知聚类个数，可以被用于本文所研究的问题。它们的不同点在于 Mean Shift 算法聚焦于利用数据点密度的峰值进行聚类，而 DBSCAN 算法则通过将足够密集的点划分为一个聚类的方式进行聚类。本文通过实验发现，DBSCAN 算法更适用于本文所研究的覆盖率度量问题，因此，本文采用 DBSCAN 算法进行求解。

DBSCAN 算法聚类过程示意图如图 4 所示，为了更好地阐述其算法逻辑，先给出如下定义^[31-33]。

1) ϵ 领域 (Epsilon neighborhood): 某点的 ϵ 领域表示所有到该点距离小于或等于 ϵ 的点集合。

2) 核心点 (core point): 若某点 ϵ 领域中的点数量大于或等于 N_{\min} ，则该点为核心点。

3) 直接密度可达 (direct density reachable): 若点 u 为核心点，点 v 在其 ϵ 领域内，则 u 到 v 直接密度可达。

4) 密度可达 (density reachable): 若点 u 和点 v 可以被多个“直接密度可达”的链路连接，则 u 与 v 之间密度可达。

5) 边界点 (boarder point): 若点 u 满足“直接密度可达”，且点 u 为非核心点，则点 u 为边界点。

6) 噪声点 (noisy point): 若点 u 既非核心点，

也非边界点，则点 u 为噪声点。

以上定义中， ϵ 和 N_{\min} 为 DBSCAN 算法的参数。其中， ϵ 决定了领域距离的大小，影响算法输出的聚类个数。 N_{\min} 决定了核心点 ϵ 领域内最少的点数，影响算法对核心点及噪声点的判断。此外，定义中点与点之间的距离体现了它们之间二维特征向量的差异性，点 $z_u = (x_u, y_u)$ 与点 $z_v = (x_v, y_v)$ 之间的距离可表示为

$$\text{dist}(z_u, z_v) = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2} \quad (7)$$

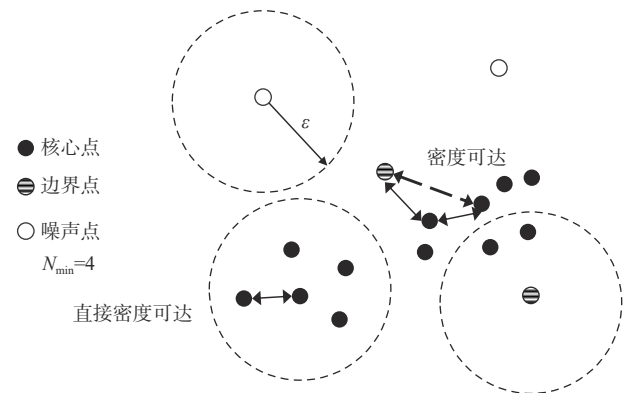


图4 DBSCAN 算法聚类过程示意图

基于以上定义，DBSCAN 算法的聚类流程如下。

步骤 1 随机选取一个初始点 u ，找到该点 ϵ 领域中的所有点，即所有到点 u 距离小于 ϵ 的点。

步骤 2 若点 u 的 ϵ 领域中包含的点数量大于或等于 N_{\min} ，将点 u 标记为核心点，根据“直接密度可达”定义，将它们（包括点 u 自身）构建为一个聚类。遍历针对点 u 所有满足“密度可达”的点，将它们纳入该聚类中并标记为核心点或边界点。

步骤 3 若点 u 的 ϵ 领域中包含的点数量小于 N_{\min} ，则随机访问其余未访问的点。

DBSCAN 算法重复步骤 2 至步骤 3，直到所有的点均被访问。最后，DBSCAN 算法将所有未聚类的点标记为噪声点，并输出聚类个数，以及每个点的聚类所属关系。

从上述步骤 1 至步骤 3 及图 4 可以看出，DBSCAN 在聚类过程中会遗留噪声点。在所考虑的室内 Wi-Fi 网络覆盖率度量问题中，所有采集的 RSSI 数据均应当被用于覆盖率的度量中。因此，需要进一步针对本文所考虑的问题改进 DBSCAN 算法，从而确立所有噪声点的聚类所属关系。为了解决上述问题，本文借鉴 K -means 算法的思想^[31-33]，通过判断各个噪声点到各个聚类中心的距离来决定噪声点的

聚类归属。针对 DBSCAN 输出的某一聚类 l ，其聚类中心在二维特征空间中的坐标可被定义为 $\mathbf{c}_l = (c_{l1}, c_{l2})$ ，其中， c_{l1} 和 c_{l2} 分别为该聚类中所有点 $\mathbf{z}_p = (x_p, y_p)$ 的第一个特征 x_p 与第二个特征 y_p 的平均值。因此， $\mathbf{c}_l = (c_{l1}, c_{l2})$ 反映了每个聚类在二维特征坐标系中的中心位置。针对任一噪声点 $\mathbf{z}_n = (x_n, y_n)$ ，本文采用如下聚类方式

$$\mathbf{z}_n \in \text{聚类 } k, \text{ 若 } \text{dist}(\mathbf{z}_n, \mathbf{c}_k) < \text{dist}(\mathbf{z}_n, \mathbf{c}_l), \forall l \neq k \quad (8)$$

基于 DBSCAN 算法并采用式(8)对噪声点进行聚类，可以得到针对服务器中存储的所有 MDT 测量报告 \mathbf{r}_p 的聚类结果。接下来可通过式(6)估计室内 Wi-Fi 网络的覆盖率，无位置坐标下基于聚类的室内 Wi-Fi 网络覆盖率度量算法如算法 1 所示。

算法 1 无位置坐标下基于聚类的室内 Wi-Fi 网络覆盖率度量算法

输入：服务器存储的所有 RSSI 数据 \mathbf{R} ，弱覆盖 RSSI 门限 δ ，DBSCAN 算法参数 ε 和 N_{\min} 。

1) 针对 \mathbf{R} 中的每个 MDT 测量报告 \mathbf{r}_p 提取特征向量 $\mathbf{z}_p = (x_p, y_p)$ ；

2) 将所有特征向量，以及参数 ε 和 N_{\min} 作为 DBSCAN 算法的输入，得到聚类个数 L ，各个特征向量的聚类归属，以及剩余噪声点；

3) 针对每个噪声点，利用式(8)进一步聚类；

4) 针对最终的聚类结果，确立 \tilde{n}_l 和 \tilde{m}_l 并根据式(6)求解覆盖率 $\tilde{\eta}$ 。

可以看出，所提覆盖率度量方法不需要每个 MDT 测量报告的位置坐标，而只基于服务器中存储的所有测量数据，可以被用于 MDT 基站回传终端 Wi-Fi 测量数据的网络覆盖测试系统。在算法复杂度方面，假设服务器中共存储了 n 个 RSSI 测量报告，由于过程 1) 需遍历所有样本，其复杂度为 $O(n)$ 。此外，过程 2) 中 n 个样本下 DBSCAN 算法的复杂度为 $O(n^2)^{[34]}$ 。最后，过程 3) 和 4) 至多需要遍历所有样本一次，复杂度均不会高于 $O(n)$ 。因此，本文所提算法的总体复杂度为 $O(n^2)$ 。

3 实验测试

为了验证本文所提覆盖率度量方法的可行性和准确性，在真实室内环境中采集 Wi-Fi 网络的 RSSI 数据，并考虑以下 3 种覆盖率度量方法进行比对。

1) 基于聚类的度量方法：本文所提方法，通过

DBSCAN 完成聚类，并应用式(8)聚类噪声点，进而采用式(6)计算覆盖率。该方法不依赖于上报数据的位置坐标，可被用于实际 MDT 回传室内 Wi-Fi 测量数据的系统中。

2) 基于位置坐标的度量方法：该方法采用式(4)，可以获取室内 Wi-Fi 网络的真实覆盖率。但需要上报数据的位置坐标，无法用于实际 MDT 回传系统。

3) 粗略度量方法：该方法将整个室内空间视为一个大栅格，并采用式(5)计算网络覆盖率，无须上报数据的位置坐标，可用于实际 MDT 回传系统。

3.1 实验环境与数据采集

本文采用华硕 AXE11000 路由器构建 Wi-Fi 网络，实验环境及测试仪表如图 5 所示。由于面向 Wi-Fi 网络的 MDT 技术还未全面商用，本文采用 zhoncent 的 WTE200V02 仪表模拟 Wi-Fi 终端获取室内各处的 RSSI。基于 3GPP 提出的 MDT 协议，将每 150 个采集的 RSSI 数据视为一个 MDT 测量报告，实验中共采集了 600 个测量报告用于覆盖率的计算。即，式(2)中服务器存储的所有 RSSI 的维度为 600×150 。值得强调的是，本文在数据采集中考虑了环境变化对 RSSI 数值的影响，因此数据采集的时间为同一天。此外，采集多天的 RSSI 数值对本文所考虑的问题意义不大。一方面，云服务器的存储能力有限，另一方面，基于多天的 RSSI 数值很难反映用户室内当前的覆盖情况。

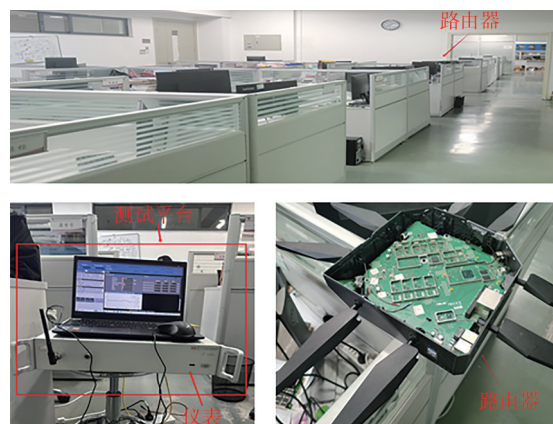


图 5 实验环境及测试仪表

由于实验中所考虑的基于位置坐标的度量方法需要确定栅格数目，因此首先研究栅格数目对该度量方法的影响。不同栅格数量 K 对基于位置坐标的度量方法的影响如图 6 所示，可以看出，当栅格数量从 4 增长到 8 时，覆盖率的度量结果出现显著变

化，因为较小的栅格可以提供更精确的度量结果。但当栅格数量继续从8增长至16时，覆盖率的度量结果几乎没有变化。这是因为当栅格面积较小时，相邻栅格的覆盖率差异不大，因此继续细化栅格不会影响覆盖率的度量结果。栅格划分情况及路由器位置如图7所示，本文将室内空间划分为8个栅格，并为测量数据打上位置坐标，用栅格标签表示。需要强调的是，栅格标签仅用于基于位置坐标的度量方法和数据呈现，在其他不需要位置坐标的覆盖率度量方法中均未考虑。

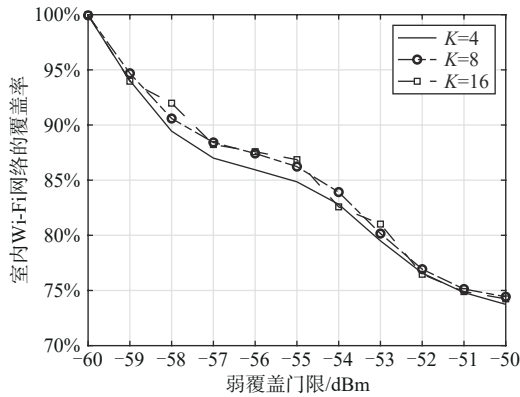


图6 不同栅格数量K对基于位置坐标的度量方法的影响

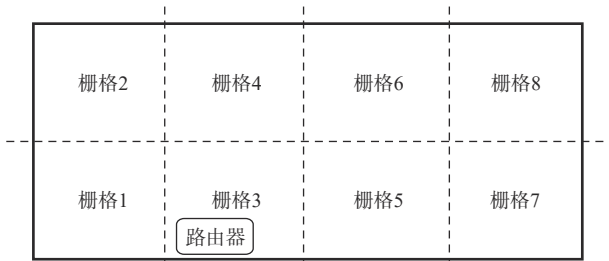


图7 栅格划分情况及路由器位置

MDT测量报告的统计特征分布如图8所示。图中每个点代表一个MDT测量报告，横轴和纵轴分别为该测量报告中150个RSSI的均值和中值，即对应特征向量 $z_p = (x_p, y_p)$ 的两个元素。为了更好地体现这两个特征的差异性，分别选取dBm和W作为单位。从图8可以看出，由于路由器位于栅格3中，该栅格中采集的RSSI均值及中值较大，相反，栅格7和8中采集的RSSI均值及中值较小。此外，所定义的均值及中值特征能够有效地区分绝大多数不同栅格上报的RSSI测量报告，但栅格1、2和4中上报的测量报告由于特征向量重叠而难以区分，在后续的实验中将发现该重叠不会显著影响本文所提覆盖率度量方法的准确度。

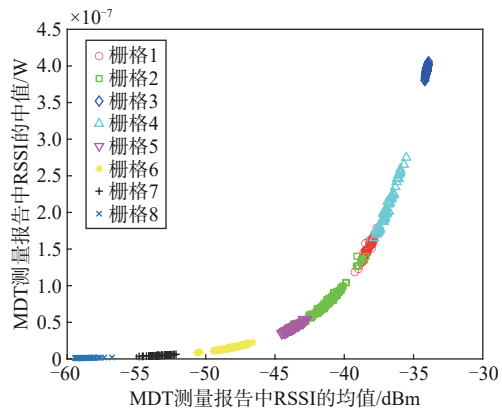


图8 MDT测量报告的统计特征分布

3.2 聚类结果

在无位置坐标条件下，根据实验中所采集的600个MDT测量报告各自的特征向量 $z_p = (x_p, y_p)$ ，采用所设计的聚类算法进行聚类。本文采用k-distance方法^[35]将DBSCAN算法的 ϵ 和 N_{min} 参数分别设置为0.2和6。原始DBSCAN算法与改进DBSCAN算法的聚类结果如图9所示。对比图8和图9中的原始DBSCAN算法聚类结果可以观察到在未知上报数据位置坐标和栅格数量的情况下，原始DBSCAN算法能够根据上报数据的特征将它们归为8类，与图8中的8个栅格相对应。然而，由于图8中栅格1、2和4的数据特征有部分重叠，图9中原始DBSCAN算法在处理这些数据的聚类时有一定偏差。最后，通过图9可以发现，原始DBSCAN算法遗留了大量未聚类的噪声点，在图中由红圈凸显，而采用本文改进的DBSCAN算法能有效地对这些噪声点进行聚类，从而验证了式(8)的可行性。

3.3 覆盖率结果

基于所有采集样本的覆盖率度量结果对比如图10所示，其中，基于位置坐标的度量方法考虑了RSSI上报数据的位置坐标，代表该室内Wi-Fi网络的真实覆盖率。从图10可以看出，所提的基于聚类的度量方法得到的覆盖率与基于位置坐标的真实覆盖率非常接近，度量准确度远高于式(5)中将整个室内空间视为一个栅格的粗略度量方法。主要原因为MDT技术无法控制用户的采集轨迹，导致一部分栅格中采集的RSSI数据较少。如果不进行聚类而直接计算覆盖率，这些数据会稀释在整个数据集中，无法准确度量覆盖率。为了更好地验证这一点，分别针对弱覆盖的栅格7和8，强覆盖的栅

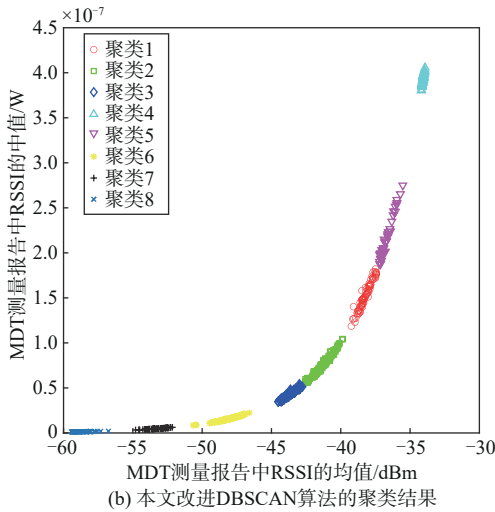
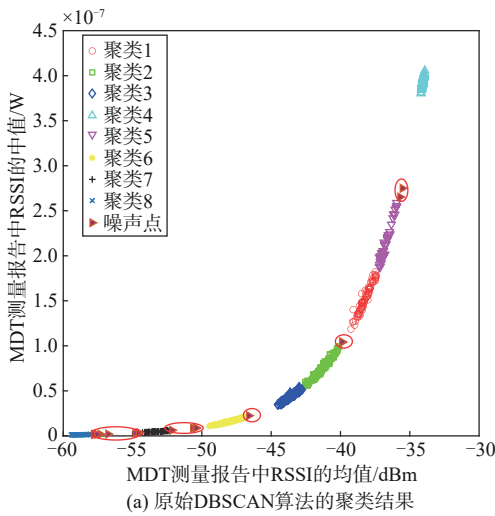


图9 原始DBSCAN算法与改进DBSCAN算法的聚类结果

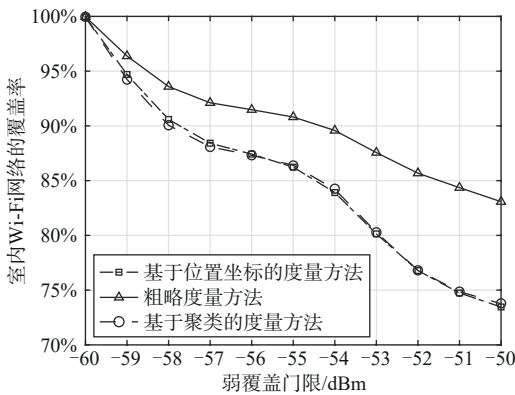


图10 基于所有采集样本的覆盖率度量结果对比

格1至6调整RSSI测量报告的数量,从而比对各个度量方法。

栅格7和8数据量减少到30%时的覆盖率度量结果对比如图11所示,将弱覆盖栅格7和8的MDT测量报告减少至30%,并采用不同度量方法评估室

内Wi-Fi网络的覆盖率。对比图10和图11可以看出数据量的改变并不影响基于位置坐标的度量方法,但能反映室内Wi-Fi网络的真实覆盖率。此外,若采用式(5)粗略度量,当弱覆盖栅格7和8中采集的测量报告减少时,度量的覆盖率将进一步偏高。栅格1至6数据量减少到20%时的覆盖率度量结果对比如图12所示,由于强覆盖栅格中的数据量减少,式(5)的计算结果将显著偏低。最后,本文所提出的基于聚类的度量方法能在不同数据量情况下保障度量准确度,且不需要位置坐标信息,为MDT基站回传终端Wi-Fi测量数据的网络覆盖测试系统提供了一种可行的覆盖率度量方法。

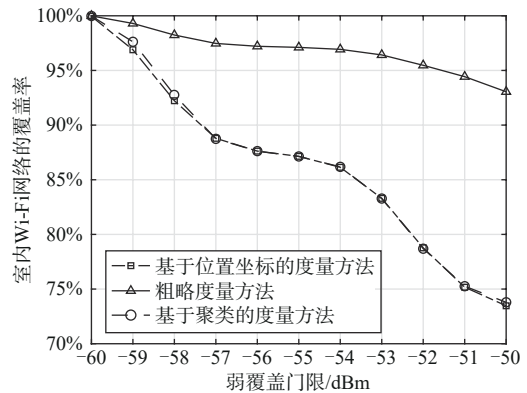


图11 栅格7和8数据量减少到30%时的覆盖率度量结果对比

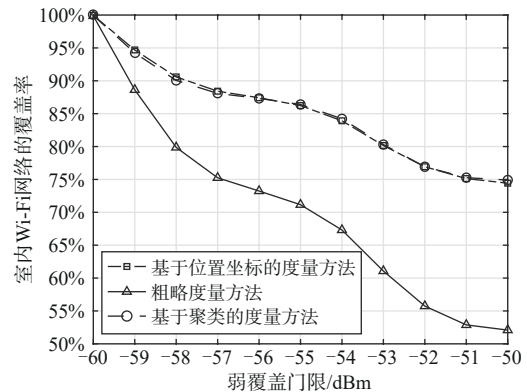


图12 栅格1至6数据量减少到20%时的覆盖率度量结果对比

4 结束语

针对当前无线网络覆盖率度量方法需要上报数据的位置坐标而室内位置坐标无法实时获取的难题,本文提出了一种基于聚类的网络覆盖率度量方法,可用于MDT基站回传终端Wi-Fi测量数据的网络覆盖测试系统。首先,本文介绍了网络系统框架并从数学角度描述了网络覆盖率的度量问题,指出

了问题求解的难点; 然后, 将覆盖率度量问题重构为一个针对 MDT 测量报告的聚类问题, 并设计相应的聚类算法进行求解, 进而给出无位置坐标下的覆盖率度量方法; 最后, 基于实验室环境真实 Wi-Fi 网络采集的 RSSI 数据进行测试, 验证所提方法的可行性与有效性, 为 MDT 的室内 Wi-Fi 网络覆盖率度量方法提供了一种新思路。

参考文献:

- [1] OUGHTON E J, LEHR W, KATSAROS K, et al. Revisiting wireless internet connectivity: 5G vs Wi-Fi 6[J]. *Telecommunications Policy*, 2021, 45(5): 102127.
- [2] NAIK G, PARK J M, ASHDOWN J, et al. Next generation Wi-Fi and 5G NR-U in the 6 GHz bands: opportunities and challenges[J]. *IEEE Access*, 2020, 8: 153027-153056.
- [3] 刘峻朋, 夏玮玮, 刘晗, 等. 面向电力业务质量保障的 NR-U 与 Wi-Fi 频谱共享[J]. *电信科学*, 2023, 39(7): 11-22.
LIU J P, XIA W W, LIU H, et al. NR-U and Wi-Fi spectrum sharing for quality guaranteeing of power services[J]. *Telecommunications Science*, 2023, 39(7): 11-22.
- [4] 黄宇, 李雨汝. Wi-Fi 6 技术现状综述[J]. *中国无线电*, 2021(5): 90-91.
HUANG Y, LI Y R. Overview of Wi-Fi 6 technology[J]. *China Radio*, 2021(5): 90-91.
- [5] SZOTT S, KOSEK-SZOTT K, GAWŁOWICZ P, et al. Wi-Fi meets ML: a survey on improving IEEE 802.11 performance with machine learning[J]. *IEEE Communications Surveys & Tutorials*, 2022, 24(3): 1843-1893.
- [6] ANDREWS J G, BACCELLI F, GANTI R K. A tractable approach to coverage and rate in cellular networks[J]. *IEEE Transactions on Communications*, 2011, 59(11): 3122-3134.
- [7] BAI T Y, HEATH R W. Coverage and rate analysis for millimeter-wave cellular networks[J]. *IEEE Transactions on Wireless Communications*, 2015, 14(2): 1100-1114.
- [8] GUO A J, HAENGGI M. Spatial stochastic models and metrics for the structure of base stations in cellular networks[J]. *IEEE Transactions on Wireless Communications*, 2013, 12(11): 5800-5812.
- [9] AKBARI I, ONIRETI O, IMRAN A, et al. Impact of inaccurate user and base station positioning on autonomous coverage estimation[C]//Proceedings of the 2015 IEEE 20th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD). Piscataway: IEEE Press, 2015: 114-118.
- [10] TAUFIQUE A, JABER M, IMRAN A, et al. Planning wireless cellular networks of future: outlook, challenges and opportunities[J]. *IEEE Access*, 2017, 5: 4821-4845.
- [11] MINOVSKI D, ÖGREN N, MITRA K, et al. Throughput prediction using machine learning in LTE and 5G networks[J]. *IEEE Transactions on Mobile Computing*, 2023, 22(3): 1825-1840.
- [12] 3GPP. 3GPP TS 37.320-Universal terrestrial radio access (UTRA) and evolved universal terrestrial radio access (E-UTRA); radio measurement collection for minimization of drive tests (MDT); Overall description (Version 10.2.0)[S]. 2011.
- [13] QURESHI H N, IMRAN A, ABU-DAYYA A. Enhanced MDT-based performance estimation for AI driven optimization in future cellular networks[J]. *IEEE Access*, 2020, 8: 161406-161426.
- [14] 潘饶元, 林宁, 刘华, 等. 一种 4G/5G 自适应覆盖评估分析方法与工具[J]. *广西通信技术*, 2022(2): 50-54.
PAN R Y, LIN N, LIU H, et al. An adaptive overcover analysis method and tool of 4G/5G network[J]. *Guangxi Communication Technology*, 2022(2): 50-54.
- [15] 杜安静, 何家辉. 基于 MDT 技术的网络综合分析方法研究与应用[J]. *电信工程技术与标准化*, 2020, 33(2): 65-68.
DU A J, HE J H. Research and application of network comprehensive analysis method based on MDT technology[J]. *Telecom Engineering Technics and Standardization*, 2020, 33(2): 65-68.
- [16] SHODAMOLA J, QURESHI H, MASOOD U, et al. Towards addressing the spatial sparsity of MDT reports to enable zero touch network automation[C]//Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2021: 1-6.
- [17] RIAZ M S, QURESHI H N, MASOOD U, et al. A hybrid deep learning-based (HYDRA) framework for multifault diagnosis using sparse MDT reports[J]. *IEEE Access*, 2022, 10: 67140-67151.
- [18] 雷泽临, 苏俭, 郭伟. 基于机器学习的蜂窝网络故障管理框架及方法综述[J]. *计算机应用研究*, 2022, 39(12): 3521-3533.
LEI Z L, SU J, GUO W. Survey of cellular network fault management framework and methods based on machine learning[J]. *Application Research of Computers*, 2022, 39(12): 3521-3533.
- [19] 王颖. 一种基于 MDT 的基站工参诊断及实现方法[J]. *长江信息通信*, 2023, 36(1): 205-208.
WANG Y. A method for base station parameter diagnosis and implementation based on MDT[J]. *Changjiang Information & Communications*, 2023, 36(1): 205-208.
- [20] 李维. 面向 5G 的覆盖效率与质量提升策略研究[J]. *信息与电脑(理论版)*, 2022, 34(14): 193-195, 220.
LI W. Research on 5G-oriented coverage efficiency and quality improvement strategies[J]. *Information & Computer*, 2022, 34(14): 193-195, 220.
- [21] DENG J, ZHENG Q B, LIU G Y, et al. A digital twin approach for self-optimization of mobile networks[C]//Proceedings of the 2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). Piscataway: IEEE Press, 2021: 1-6.
- [22] SKOCAJ M, AMOROSA L M, GHINAMO G, et al. Cellular network capacity and coverage enhancement with MDT data and deep reinforcement learning[J]. *Computer Communications*, 2022,

195: 403-415.

- [23] SÁNCHEZ-MARTÍN J M, TORIL M, WILLE V, et al. On the improvement of cellular coverage maps by filtering MDT measurements[J]. IEEE Transactions on Mobile Computing, 2023, 22(7): 4119-4133.
- [24] 3GPP. 3GPP TS 38.331-NR; Radio Resource Control (RRC); Protocol specification (Version 16.7.0)[S]. 2021.
- [25] SINGH N, CHOE S, PUNMIYA R. Machine learning based indoor localization using Wi-Fi RSSI fingerprints: an overview[J]. IEEE Access, 1083, 9: 127150-127174.
- [26] ZHOU M, LI Y H, TAHIR M J, et al. Integrated statistical test of signal distributions and access point contributions for Wi-Fi indoor localization[J]. IEEE Transactions on Vehicular Technology, 2021, 70(5): 5057-5070.
- [27] LIU S J, CHANG R Y, CHIEN F T. Analysis and visualization of deep neural networks in device-free Wi-Fi indoor localization[J]. IEEE Access, 2019, 7: 69379-69392.
- [28] TONG X Y, WANG H, LIU X L, et al. MapFi: autonomous mapping of Wi-Fi infrastructure for indoor localization[J]. IEEE Transactions on Mobile Computing, 2023, 22(3): 1566-1580.
- [29] LAM K H, CHEUNG C C, LEE W C. RSSI-based LoRa localization systems for large-scale indoor and outdoor environments[J]. IEEE Transactions on Vehicular Technology, 2019, 68(12): 11778-11791.
- [30] 3GPP. 3GPP TS 32.422 - Telecommunication management; Subscriber and equipment trace; Trace control and configuration management (Version 18.0.0)[S]. 2023.
- [31] CHEN Y W, ZHOU L D, PEI S W, et al. KNN-BLOCK DBSCAN: fast clustering for large-scale data[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(6): 3939-3953.
- [32] DONG G S, JIN Y, WANG S W, et al. DB-kmeans: an intrusion detection algorithm based on DBSCAN and K-means[C]//Proceedings of the 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS). Piscataway: IEEE Press, 2019: 1-4.
- [33] LIN Y, GIACOUMIDIS E, O'DUILL S, et al. DBSCAN-based clustering for nonlinearity induced penalty reduction in wavelength conversion systems[J]. IEEE Photonics Technology Letters, 2019, 31(21): 1709-1712.
- [34] JANG J, JIANG H. DBSCAN++: towards fast and scalable density clustering[C]//Proceedings of the 36th International Conference on Machine Learning. New York: Curran Associates, 2019: 3019-3029.
- [35] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. CA: AAAI Press, 1996: 226-231.

[作者简介]



谢泽锋(1983-), 男, 现就职于中国移动通信集团广东有限公司, 主要研究方向为业务质量和客户感知分析优化。



陈伟栋(1974-), 男, 中国移动通信集团广东有限公司高级工程师, 主要研究方向为业务质量和客户感知分析优化。



黄黎霞(1999-), 女, 深圳大学电子与信息工程学院硕士生, 主要研究方向为无线通信系统、射频系统校准与测量、机器学习等。



顾一帆(1990-), 男, 博士, 深圳大学电子与信息工程学院助理教授, 主要研究方向为无线通信系统、射频系统校准与测量、信息年龄和图神经网络。



张博钧(1981-), 男, 中国信息通信研究院南方分院(深圳信息通信研究院)副院长, 主要研究方向为无线通信设备检测技术研究与标准研制、国内外认证政策与技术。



全智(1978-), 男, 博士, 深圳大学电子与信息工程学院特聘教授, 主要研究方向为无线通信系统、射频系统校准与测量、数据驱动信号处理和机器学习。