

RW-MC: 基于随机游走的自适应矩阵填充算法

王新恒, 王倩云, 王佳杰, 赵国锋, 靳文强

(重庆邮电大学通信学院, 重庆 400065)

摘 要: 为了对软件定义无线网络系统中虚拟接入点 (VAP) 状态信息进行实时测量, 根据实际网络中虚拟接入点性能的数据特征, 提出一种基于随机游走的自适应矩阵填充算法 (RW-MC)。首先, 基于离散度和覆盖度的采样模型确定初始样本点; 然后, 利用随机游走模型对之前时隙的采样点序列建模分析, 确定新时隙的测量点; 最后, 比较相邻窗口的恢复矩阵中重叠部分的误差率与标准误差, 实现测量点的动态自适应。实验表明, 该测量方法能够在低采样率、低误差的情况下实现对全网 VAP 实时感知。

关键词: SDWN; 矩阵填充; RW-MC; 随机游走

中图分类号: TP319

文献标识码: A

RW-MC: self-adaptive random walk based matrix completion algorithm

WANG Xin-heng, WANG Qian-yun, WANG Jia-jie, ZHAO Guo-feng, JIN Wen-qiang

(School of Communication, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Concerning the continually perceiving performance of virtual access points (VAP) was urgent in software-defined wireless network (SDWN), with the features of VAPs' measurement data (VMD), a self-adaptive matrix completion algorithm based on random walk was proposed, named RW-MC. Firstly, the discrete ratio and covering ratio of VMD account for a sample determination model was used to claim initial samples. Secondly, random walk model was implemented for generating sampling data points in the next iteration. Finally, a self-adaptive sampling redress model concerning the differences between the current error rates and normalize error rates of neighboring completion matrices. The experiments show that the approach can collect the real-time sensory data, meanwhile, maintain a relatively low error rate for a small sampling rate.

Key words: SDWN, matrix completion, RW-MC, random walk

1 引言

近年来, 随着智能终端的快速普及, 全球移动数据业务呈指数式增长。软件定义无线网络 (SDWN, software defined wireless network)^[1]作为 5G^[2,3]关键技术之一, 解决了传统无线网络中架构封闭、僵化的缺陷, 实现了控制与转发分离, 满足了移动终端的无缝切换、安全认证等需求。基于

SDN 和 NFV 理论, 在同一接入点 (AP, access point) 设备上的不同虚拟 AP (简称 VAP) 创建相互独立的无线网络, 实现终端无线接入^[4]。由于 VAP 具有独立的 BSSID, 并从 MAC 层模拟物理 AP 的功能。因此, 每个物理 AP 产生的多个 VAP 会平分带宽资源和信道速率。为了保证用户终端的服务质量, 本文每个物理 AP 虚拟出 6 个 VAP。控制器要实现系统负载均衡及调度控制, 就需要对 VAP 性能进行实

收稿日期: 2016-12-10; 修回日期: 2017-06-27

通信作者: 赵国锋, zhaogf@cqupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61402065); 国家重点基础研究发展计划 (“973” 计划) 基金资助项目 (No.2012CB315803, No.2012CB315806)

Foundation Items: The National Natural Science Foundation of China (No.61402065), The National Basic Research Program of China (973 Program) (No.2012CB315803, No.2012CB315806)

时感知。大量的研究工作^[5-8]将 VAP 信号强度、VAP 关联用户数和 VAP 数据流量 3 个性能指标作为控制器决策的重要依据。因此,运营商在进行无线网络资源优化管理时,对 VAP 的上述 3 个性能指标进行有效且准确的测试具有十分重要的意义。然而,对系统中所有 VAP 节点的性能数据均采用直接测量的方式,获取资源调度执行所需的性能参数会存在诸多方面的问题。首先,SDWN 中 VAP 节点规模庞大,对每个 VAP 节点性能数据都进行直接测量必然耗费高昂的成本;其次,大规模的数据测量不仅会消耗系统存储资源和处理资源,在性能数据的测量过程中,其测试命令的传输还会占用额外的网络带宽资源,从而加重网络负载,影响终端用户的通信质量。因此,提出一种低成本、高准确率的全网 VAP 性能数据测量方法迫在眉睫。

压缩感知作为一种新的采样理论,利用信号的稀疏特性,在远小于 Nyquist 采样率的条件下,根据随机采样获取信号的离散样本,通过非线性重建算法恢复丢失信号。文献[9]利用基于随机游走和压缩感知的方法对强连通以及低节点度数的城市道路进行分析,设计了一种数据恢复算法,仅用少量探测车推测出交通全网络的道路延时和拥塞情况。为了实现细粒度的流量测量, Malboubi 等^[10]提出了一种基于 SDN 的测量架构 iSTAMP,该架构利用压缩感知理论设计最佳二进制测量矩阵,同时提出了一种设计有效的压缩流量聚集矩阵的方法,实现低成本、高准确率的测量方案。基于压缩感知理论的测量方法虽然可以节省测量开销,但大多数应用程序在实际情况下并没有明显的稀疏性特征。

矩阵填充技术作为一个引人注目的新领域,是压缩感知理论向二维空间的扩展与衍生。研究表明低秩矩阵可以通过合适的算法进行精确的重构^[11]。Gong^[12]利用 SDN 的灵活性设计出一个智能的测量架构——SNIPER,采用矩阵填充技术和优化算法实现对全网性能的准确监测。Xie^[13]将矩阵填充技术应用到端到端网络中进行网络性能监控,提出了一种基于矩阵填充的自适应连续采样算法。He^[14]和 Cheng^[15]针对无线传感器网络中因资源有限导致的数据采集困难问题,利用采集数据在空间和时间上的稀疏特性,提出了一种结合矩阵填充和稀疏约束的数据恢复算法。文献[16]利用无线传感网络收集天气信息,研究发现该数据集具有低秩、相对稳定的特性,提出基于矩阵填充的线上数据收集方

法,实现高准确率、低成本的全网数据测量。因此,本文将探索如何将矩阵填充技术有效引入到 SDWN 网络测量中,以实现全网 VAP 的低成本、高准确率的检测。

现有的矩阵填充解决方案通常假设数据矩阵有一个已知且固定的低秩,文献[17]指出采样数目 m 必须满足 $m \geq Cn^{\frac{6}{5}}r\ln n$ 条件,才能以接近 1 的概率恢复出该矩阵的所有数据。然而,无线网络的不稳定性使 VAP 性能动态变化,测量数据矩阵的秩随时间变化不再是一个固定值。为了研究 VAP 的网络性能特征,对 90 个 VAP 设备进行连续数据采集,发现 VAP 性能数据矩阵的秩确实存在随时间动态改变的问题。根据矩阵填充理论,随着数据矩阵秩的增大,需要更多采样点才能实现数据准确恢复。因此,若想实时恢复 VAP 动态变化的性能数据,则需以自适应的方式调整每一时隙的采样数目。

实现实时矩阵恢复,确定新时隙的采样个数以及采样点的选择同样重要,而目前的多数研究缺乏对采样点位置选择的考虑。随机游走作为一种解决数据预测的重要方法,广泛应用于互联网及金融股票市场。文献[18]提出一种联合考虑用户和产品的推荐算法。根据用户间以及用户与产品的关系,提出一种主题模型,建立用户—产品的二分图,利用随机游走测试它们的相关性,设计了用户和产品的综合推荐系统。文献[19]利用随机游走推测源节点(发数据分组)的位置以保护源节点。监测网络的边缘部分,游走得到的第一个边缘节点分布反映出源节点的相关位置,据此实现攻击监测。可以看出,随机游走是通过对原始数据集的学习,并根据当前状态以最大概率预测出下一状态。将随机游走用于系统性能数据的测量,在高准确率恢复采样的方式下,发现采样点间存在的规律,借助该规律构造随机游走图和转移概率矩阵,以迭代收敛的方式科学地筛选出下一时隙的最优采样节点。经过若干次随机游走,得到最优采样序列,为提高新时隙的数据恢复准确率起到了重要作用。

本文提出了一种基于随机游走的矩阵填充算法——RW-MC 算法,该方法以较低的测量样本率实现全网 VAP 性能测量。主要工作如下: 1) 基于对大量 VAP 网络性能的分析,发现测量数据具有低秩性、时间稳定性和相邻时隙数据稳定性和秩的相对稳定性的特征,利用这些特性引入矩阵填充理

论,即通过测量部分 VAP 的性能以恢复全网的网络状态,从而减少 VAP 测量点;2) 利用随机游走模型对之前时隙的采样节点序列进行建模分析,预测下一时隙的采样节点序列,如此,在新时隙有一个更科学合理的采样方式,最终通过对新时隙采样点的实际测量,应用 SVT 算法以较低地恢复误差实现进行整体数据重构;3) 为最大化地减少测量样本点,本文将相邻窗口所得到的恢复矩阵中交叉部分的误差率与基准重构误差进行比较,实现自适应调整测量样本数。

2 问题分析及研究思路

2.1 测量问题

本文引入矩阵填充理论,利用直接测量得到的少量 AP 性能参数近似准确地恢复出全网 AP 的性能参数,从而降低网络开销,提高系统稳定性。

假设 SDWN 网络中 VAP 节点数为 N , 测量时隙为 T , 则定义 $\mathbf{X}_{N \times T}$ 为测量矩阵, VAP 为矩阵的行, 时隙为矩阵的列, 其中, X_m 表示矩阵 \mathbf{X} 中第 n 个 VAP 在时隙 t 内的性能数据值。

定义 1 二进制抽样矩阵 $\mathbf{B}_{N \times T}$ 的表达如式(1)所示。

$$\mathbf{B}_{N \times T} = (\mathbf{B}_m)_{N \times T} = \begin{cases} 1, \text{直接测量} \\ 0, \text{其他} \end{cases} \quad (1)$$

其中, 矩阵 \mathbf{B} 中元素值为 1 表示该节点为直接测量点, 元素值为 0 表示该节点为无需直接测量。

定义 2 直接测量矩阵 $\mathbf{M}_{N \times T}$ 的表达如式(2)所示。

$$\mathbf{M}_{N \times T} = \mathbf{X}_{N \times T} \bullet \mathbf{B}_{N \times T} \quad (2)$$

其中, \bullet 运算符表示 $M_m = X_m \times B_m$; 矩阵 \mathbf{B} 为二进制抽样矩阵, 其中, 1 对应的下标即为直接测量点集合。由于只对部分节点直接测量, 因此, 矩阵 \mathbf{M} 中会存在空缺元素, 为此, 将未测量的元素值标记为 0。

相关研究表明^[12,18], 测量样本数满足式(3)时, 才能以接近 1 的概率恢复出原矩阵, 其中, C 是常数, n 为矩阵的阶数, r 为矩阵的秩, m 为测量样本数。但是在实际测量中, 随着时隙的增加需要对性能数据进行实时测量, 并且测量执行之前是无法预知结果矩阵秩的大小。

$$m \geq Cn^{\frac{6}{5}}r\ln n \quad (3)$$

因此, 以下 2 个问题是本文研究重点。

- 1) 确定 t 时隙采样个数 m_t ;
- 2) 确定 t 时隙采样节点序列 B_t 。

2.2 矩阵秩的特征分析

从式(3)可以看出, 测量样本数 m 与矩阵秩 r 有关, 而数据矩阵的秩会随着时隙的移动发生变化, 使新时隙滑动窗口进行矩阵填充变得更加困难。变化的秩意味着需要动态地调整采样点数目才能完成高准确率矩阵填充, 而在实际应用中, 并不知道新时隙滑动窗口矩阵的秩。Xie^[13]提出低秩性、相邻时隙稳定性、秩相对稳定性是使用新时隙测量点精确恢复矩阵的必要条件。

本文采用 SDWN 实验平台测量的真实数据为仿真数据集。SDWN 实验平台由 1 台控制器 (操作系统为 CentOS6.3 CPU 为奔腾 E5300, 双核, 2.60 GHz 内存 3 GB)、1 台 SDN 交换机 (比威 S6228SP 24 端口, 千兆带宽) 以及 15 个 Netgear WNDR3800 AP 设备组成。综合考虑终端需求的带宽大小和物理 AP 本身的资源情况, 在 SDWN 实验床中的每个 AP 设备均添加 6 个 VAP, 共 90 个 VAP 节点。在测量过程中每隔 5 s 对所有 VAP 执行一次测量, 测量指标包括 VAP 信号强度、已关联用户数和数据流量, 持续测试 12 h。为了保证测量数据的真实性, 本文选择了某工作日的 9:00~17:00 时间段, 在重庆邮电大学逸夫实验楼 3 楼进行数据采集。为了降低异常数据对算法仿真结果的干扰, 本文每 30 s 取一次数据平均值作为一个时隙的数据读取值, 将计算后的值表示为矩阵的形式, 其中, VAP 为矩阵的行, 时隙为矩阵的列。

定义矩阵 $\mathbf{S}_{N \times T}(t)$ 为 t 时间点测量得到的 VAP 性能数据矩阵, 矩阵元素 S_{ij} 表示 VAP _{i} 在第 j 个时隙下的性能数据值, 矩阵 \mathbf{S} 中每一列表示对应时段下的每个 VAP 节点的性能数据, 矩阵 \mathbf{S} 中每一行表示单个 VAP 节点在不同时隙下的性能数据。

1) 低秩性

图 1 中 X 轴表示按照降序排列的前 k 个矩阵奇

异值; Y 轴 $g(k) = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{i=1}^r \sigma_i^2}$, 其中, σ_i 表示第 i 个奇异

值。从图 1 中可知, 前 10% 的奇异值占奇异值总和的百分比在 80% 左右。因此, 可以得出以下结论: 这 3 种测试数据在该实验环境中都具有明显的低秩特性。

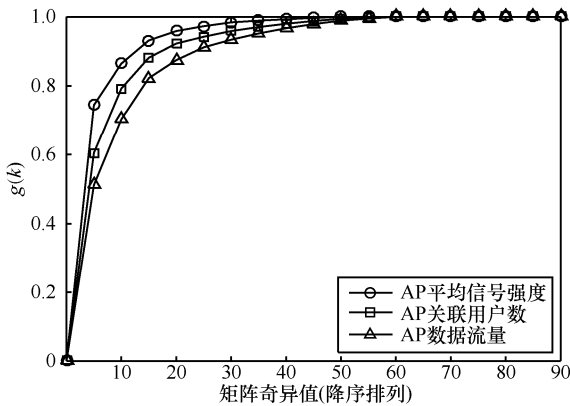


图 1 矩阵奇异值 $g(k)$

根据 VAP 的工作原理,通过虚拟化技术使所有 VAP 利用同一无线射频模块发射信号,且均处于同一无线信道中,并平分物理 AP 的带宽资源。因此,处于某一物理环境下,单个物理 AP 设备下的所有 VAP 的性能参数具备相似性,如关联用户、数据流量的近似平均分布。

2) 相邻时隙数据稳定性

图 2 中 X 轴表示 2 个连续时隙之间的归一化差值;Y 轴表示累积概率。观察到通过归一化处理过的相邻时隙数据差小于 0.1 的超过 90%。这些结果表明相邻时隙内的 AP 性能数据变化平稳。

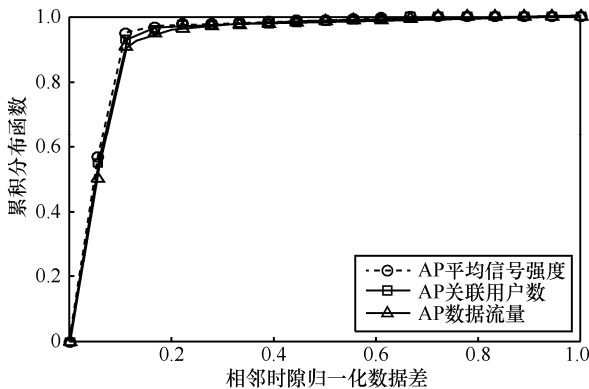


图 2 相邻时隙数据稳定性

3) 秩的相对稳定性

图 3 中 X 轴表示随着时间变化每个滑动窗口内矩阵的初始时隙,设置滑动窗口大小为 240 时隙,而整个测试一共有 1440 个时隙,所以 X 轴的范围为 0~1 200, Y 轴为对应矩阵的秩。数据矩阵的秩是动态变化的,但相邻矩阵的秩变化不大。因此,在已知矩阵中的部分元素时,利用矩阵填充理论能有效重构缺失的元素。

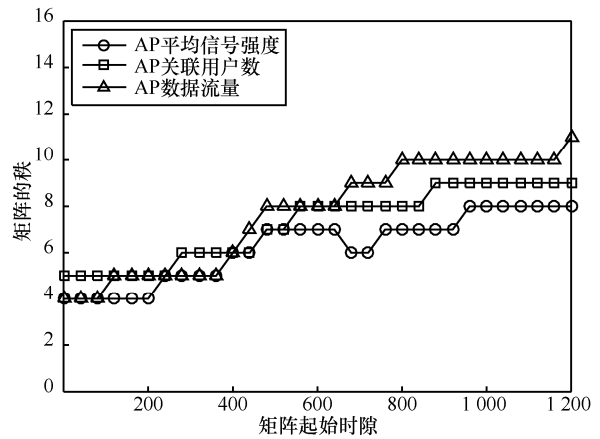


图 3 随时间变化的矩阵秩特征

2.3 研究思路

为了以较低的测量成本实现全网 VAP 数据流量的测量,本文提出了基于滑动窗口模型的 RW-MC 方法。RW-MC 的滑动窗口以时隙为单位,矩阵的行为 VAP 节点,矩阵的列为时隙。滑动窗口大小固定并随着时间变化向右移动,定义包含当前时隙的窗口为活跃窗口。

如图 4 所示,假设测试中包含 2 个物理 AP,即 12 个 VAP 节点,其中,滑动窗口大小设定 $T=7$ 。图中包含 2 个窗口,第 1 个窗口是从时隙 3 到时隙 9,记为 $X_{12 \times 7}(9)$;第 2 个窗口是从时隙 4 到时隙 10,记为 $X_{12 \times 7}(10)$ 。由于当前时隙为 10,即第 2 个窗口为当前活跃窗口。从图 4 可以看出,相邻的 2 个窗口矩阵仅有一列元素不同,为了研究前后 2 个矩阵的秩之间的关系,下面介绍 2 个定理。

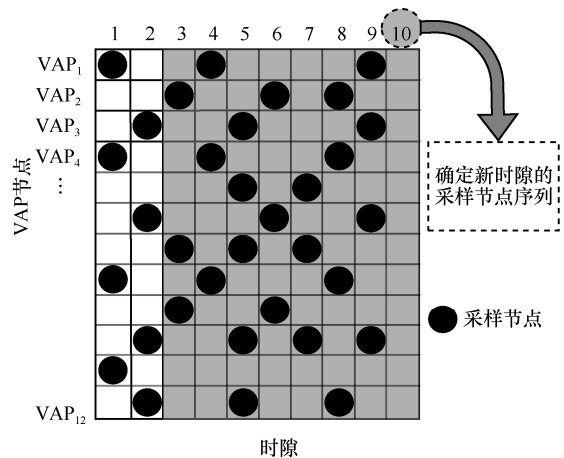


图 4 滑动窗口模型

定理 1 给定矩阵 $A \in R^{m \times n}$, 矩阵 $B \in R^{m \times n}$, $(A, B) \in R^{m \times 2n}$ 为矩阵 A 的增广矩阵,则矩阵秩满足以下关系。

$$\begin{cases} \max\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\} \leq \text{rank}(\mathbf{A}, \mathbf{B}) \\ \text{rank}(\mathbf{A}, \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \end{cases} \quad (4)$$

其中, $\text{rank}()$ 表示求解矩阵秩函数。

若矩阵 \mathbf{B} 为一个非零向量 $\mathbf{B} \in R^m$, 则矩阵秩满足如式(5)的关系。

$$\text{rank}(\mathbf{A}) \leq \text{rank}(\mathbf{A}, \mathbf{B}) \leq \text{rank}(\mathbf{A}) + 1 \quad (5)$$

定理 2 已知 VAP 数据流量窗口矩阵 $\mathbf{X}_{N \times T}(t)$ 和 $\mathbf{X}_{N \times T}(t+1)$, 其中, $\text{rank}(\mathbf{X}_{N \times T}(t)) = r$, 则矩阵 $\mathbf{X}_{N \times T}(t+1)$ 的秩满足式(6)。

$$r - 1 \leq \text{rank}(\mathbf{X}_{N \times T}(t+1)) \leq r + 1 \quad (6)$$

上述 2 个定理体现出相邻窗口矩阵的秩具有稳定性, 它将作为新时隙数据采样的停止条件。文献[13]提出 3 种学习采样的方法, 即每个时隙的采样点个数和位置都是固定的, 这种方式缺乏对采样过程中 VAP 节点间关系的考虑。因此, 本文提出基于随机游走的自适应采样算法。

1) 根据测试数据集确定初始随机游走图和基准重构误差。将初始滑动窗口作为数据集, 并对该窗口内所有 VAP 的数据进行直接测量, 得到数据矩阵。结合采样式与基于离散度、覆盖度的采样修正模型得到滑动窗口内的采样节点序列。根据采样节点的数据使用 SVT 算法对整个初始滑动窗口数据进行恢复, 与真实数据比较得到基准重构误差, 并以采样节点序列构造初始随机游走图。

2) 利用随机游走模型对之前时隙的采样点序列建模分析, 确定新时隙的测量点。随着滑动窗口的右移, 通过随机游走过程实时推测下一时隙最优采样点位置并更新随机游走图。

3) 将相邻窗口的恢复矩阵中重叠部分的误差率与标准误差比较, 实现测量点的动态自适应。在滑动窗口随时间右移的过程中, 首先保证相邻滑动窗口内采样数一致, 进而基于相邻滑动窗口重构矩阵重叠区域误差与基准重构误差的对比结果动态增减, 实现每一时隙采样数目自适应, 最终得到合理的新时隙采样点序列。对新时隙采样点进行实际测量, 通过 SVT 算法恢复出滑动窗口内全部数据, 由此便推测出新时隙内所有 VAP 节点的有效性数据。

3 随机游走模型

3.1 构造 VAP 节点随机游走图

通过初始窗口数据矩阵的采样节点序列构造出 VAP 节点的初始随机游走图, 如图 5 所示。假设

SDWN 中有 N 个 VAP 节点, 且分别用 1 到 N 的数字进行标记, 记 VAP 节点集合为 $N = \{1, 2, 3, \dots, N\}$ 。

滑动窗口大小为 T , 测量得到的 T 个时隙的采样节点序列作为训练集合 $D = \{D_1, D_2, D_3, \dots, D_T\}$, 其中, $D_i = \{x_{n+1}, x_{n+2}, x_{n+3}, \dots, x_{n+d}\}$, $x_i \in N$ 为第 i 次采样对应的 VAP 节点。首先, 通过节点集合 N 和训练集合 D 构造初始化随机游走图, 然后利用随机游走模型通过点与点之间的连通性刻画各个 VAP 节点之间的关联性。其基本思路为将集合 N 中的每个 VAP 节点 $i \in N$ 映射为图中的一个点 v_i , 若采样序列 D 中存在节点 $i \in N$ 到节点 $j \in N$ 的跳转 $i \rightarrow j (i \neq j)$, 则将顶点 v_i 和 v_j 连接, 记为 (v_i, v_j) 。由 VAP 节点集合 N 和训练集合 D 导出的随机游走图记为式(7)~式(9)。

$$G = (V, E) \quad (7)$$

$$V = \{v_i | i \in N, 1 \leq i \leq k\} \quad (8)$$

$$E = \{(v_i, v_j) | v_i, v_j \in V, i \neq j\} \quad (9)$$

其中, V 为图中的顶点集合, 其元素对应节点集合中的一个 VAP 节点, E 为图中的边, 其元素对应着 VAP 节点之间的跳转情况, (v_i, v_j) 表示在图 G 中存在一条 v_i 到 v_j 的边。

然后, 计算随机游走图 G 的权重矩阵 \mathbf{W} , 如式(10), 边的权值为训练集合 D 中 VAP 节点之间的转移次数。

$$W_{ij} = B(v_i, v_j), (v_i, v_j) \in E \quad (10)$$

其中, $B(v_i, v_j)$ 为训练集合 D 中的所有时隙节点序列中节点 i 到节点 j 的转移次数。

根据随机游走图 G 的权重矩阵 \mathbf{W} 计算转移概率矩阵 \mathbf{P} , 如式(11)所示。其中, P_{ij} 表示顶点 v_i 转移到顶点 v_j 的转移概率。

$$P_{ij} = \frac{W_{ij}}{\sum_i W_{ij}} \quad (11)$$

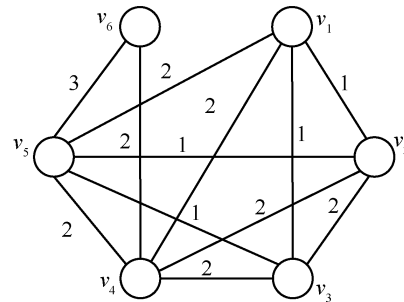


图 5 随机游走图的构造示例

3.2 随机游走过程

建立完初始随机游走图后, 就可以通过随机游走过程找到下一时隙的最佳采样点。以第 t 个时隙最后一个采样点为起点依据随机游走图 G' 实现随机游走过程, 得到稳定状态的概率分布向量 Π , 用于预测第 $t+1$ 个时隙的首个采样节点。

随机游走过程包括 4 个输入参数: 邻接矩阵 p 、初始概率分布向量 s_0 、跳转发生概率 α 、发生跳转时跳转到图中每个顶点的概率分布向量 d 。每次游走过程后的输出概率分布向量记作 s , 则 s 的计算方法为

$$s = (1 - \alpha)s_0 + \alpha \cdot d \quad (12)$$

将向量 s 作为式(12)的输入 s_0 , 反复迭代式(12)直至收敛, 将此时的概率分布向量记作 Π , 向量 Π 即为稳定状态的概率分布向量。

为了应用式(12), 首先利用训练集合 D 计算得到随机游走图 G 的权重矩阵 W 和转移概率矩阵 P , 此时, 将训练集合 D 训练得到的转移概率矩阵 P 用于新时隙采样节点序列的预测中, 在随机游走过程中, 粒子以转移概率矩阵 P 随机游走到其他顶点。因此, 邻接矩阵 p 为训练集合 D 的转移概率矩阵 P 。

根据训练得到的 VAP 节点转移概率矩阵 P , 得到 x_n 发生跳转时跳转到图中每个顶点的概率分布向量, 即

$$d = P(x_n) \quad (13)$$

初始概率分布向量 s_0 则由随机游走图 G' 得到, 为图 G' 中从 x_n 转移到图中其他顶点的概率分布。

$$s_0 = \left\{ \frac{W'_{x_n 1}}{\sum_i W'_{x_n i}}, \frac{W'_{x_n 2}}{\sum_i W'_{x_n i}}, \dots, \frac{W'_{x_n M}}{\sum_i W'_{x_n i}} \right\} \quad (14)$$

其中, $W'_{x_n j}$ 表示顶点 x_n 转移到顶点 x_j 的次数, α 的取值一般为 0.15。

将式(12)中所需的 4 个输入参数代入, 可以得到概率分布向量 s , 经过反复迭代式(12)直至其收敛。为此, 可得出稳定状态的概率分布向量 Π , Π 表示在稳定状态下每个顶点的概率分布。

4 基于随机游走的矩阵填充算法

4.1 确定初始采样样本

设全网 VAP 节点数为 N , 滑动窗口大小为 T , 记窗口矩阵为 $X_{N \times T}(t)$ 。在初始化阶段, 所有 VAP

节点执行测量任务, 持续 T 个时隙, 并将测量数据上传给控制器, 记为 $X_{N \times T}(T)$, 且该窗口矩阵的秩 $r = \text{rank}(X_{N \times T}(T))$ 。因此, 根据式(3)可计算出该窗口矩阵的采样数为 m 。

研究表明, VAP 节点直接测量的选择对于矩阵的重构正确率有很大的影响。文献[20]描述了 2 种采样模型: 伯努利模型和均匀模型。在伯努利模型中, 每个节点被选为采样节点的概率 $p = \frac{m}{n_1 n_2}$, 其中, n_1 和 n_2 分别表示矩阵的行数和列数。在均匀采样模型中, 随机从矩阵中选取 Ω 个节点作为采样节点。然而这 2 种采样模型并不能保证采样节点具有较大的离散度, 同时也不能够确定矩阵的各行各列都至少被一个采样点覆盖, 这都会对矩阵重构的准确率产生影响。

本文针对上述 2 种采样方式所出现的问题, 进行了以下改进并提出了基于离散度和覆盖度的采样模型。

1) 从初始滑动窗口的数据矩阵 $X_{N \times T}$ 中随机选取 m 个样本点。

2) 记每个样本点为 X_n , n 为 AP 的编号, t 为时隙, 计算全部样本点 n 和 t 的标准差 σ_n 和 σ_t 。

$$\sigma_n = \frac{\sum_i (n_i - \bar{n})^2}{m}$$

$$\sigma_t = \frac{\sum_i (t_i - \bar{t})^2}{m}$$

3) if $\sigma_n \geq \text{tau}_n \&\& \sigma_t \geq \text{tau}_t$ then

4) 采样点具有较高的离散度;

5) else

6) 采样点离散度较低, 重新采样;

7) end if

8) if $\text{sum}(AP_i) \neq 0 \&\& \text{sum}(t_j) \neq 0$ then

9) 所有行和列都有数据被采集到;

10) else

11) 有行或列存在没有被采样点覆盖的情况, 重新采样;

12) end if

13) 得到一个有效的采样结果, 集合为 D ;

14) 通过 SVT 算法重构, 得到数据矩阵 \hat{X} ;

15) 对比 $X_{N \times T}$ 与 $\hat{X}_{N \times T}$, 得到基准重构误差率 ε_0 ;

$$\varepsilon_0 = \frac{\sum |X_{ij} - \hat{X}_{ij}|}{\sum |X_{ij}|}$$

其中, tua_n 、 tua_t 分别为 m 个采样点 AP 编号和时隙编号标准差的阈值, 这 2 个值与数据矩阵的大小有关, 其值通过多次的训练来确定。当一次采样结果离散度满足步骤 2) 与步骤 3) 时, 这些采样点有较高的离散度; 如果不满足, 则重新进行采样。

当拥有较高离散度之后, 继续验证矩阵的各行各列都被至少一个采样点覆盖。如不满足此条件, 则重新进行采样; 如满足, 则是一个可靠有效的采样结果。

4.2 基于随机游走模型的自适应采样阶段

由于终端用户的上网行为具有规律性, 第 t 时隙的采样节点序列与其之前时隙的采样序列存在相关性, 因此可以根据第 t 时隙之前的采样节点序列呈现出的规律特性, 利用随机游走过程对第 t 个时隙的采样节点序列进行预测。本文所提出的预测方法是利用随机游走模型对之前时隙的采样节点序列进行建模分析, 预测下一时隙的采样节点序列。如此在新时隙有一个更科学合理的采样方式, 以获得更低的恢复误差。

本文运用基于随机游走模型的时隙采样节点序列预测算法, 根据一段时间的采样节点序列构造随机游走图, 其顶点数即为系统中的 VAP 数量且 VAP 节点与顶点一一对应, 假设各顶点上都存在一个可移动粒子, 根据各粒子历史运动轨迹获得训练集合 D , 然后每个顶点走到其他顶点的转移概率可以通过 D 计算得到, 最后建立转移概率矩阵。假设以随机游走图中某一项点上的粒子 x_i 为本次游走过程的起始点, 它经过 n 步转移到达另一顶点 x_j , 根据转移概率矩阵得到其转移概率为 p_{ij} 。由马尔可夫过程性质可以获得, 当相邻概率值相等或接近相等时, 认为该随机游走过程收敛到一个平稳状态。

由 3.2 节可知, 通过构造随机游走图并据图进行随机游走, 最终可以根据稳定状态下的概率分布向量 Π 得到从 x_n 出发转移到达哪个顶点的概率最大, 从而预测出顶点 x_n 的下一个顶点 x_{n+1} 。

$$x_{n+1} = \text{avg} \max_{1 \leq j \leq M} \{\pi(j)\} \quad (15)$$

当预测出 x_{n+1} 后, 则更新随机游走图 G' 。 x_n 对应的顶点记为 v_i , x_{n+1} 对应的顶点记为 v_j , 改变随机游走图 G 中边的权值, 将边 (v_i, v_j) 的权值加 1。接下来, 继续以 x_{n+1} 为起点根据随机游走图 G' 实现

随机游走过程, 预测下一个采样节点, 以此类推, 直到预测出第 $k+1$ 时隙的最后一个采样节点 x_{n+d} 为止。

通过随机游走, 可以清楚地得到每一次走后的采样节点, 但是每一个时隙当中采样点数目仍无法确定。为了解决这个问题, 下面, 提出了基于随机游走的自适应连续采样算法。

1) 窗口固定采样

由图 3 和 2.3 节可以看出, 在滑动窗口移动过程中, 由于相邻数据矩阵的秩之间具有稳定性, 则保持窗口内采样节点数不变, 使新时隙采样数目与移出滑动窗口的时隙保持一致, 如图 6 所示, $t+1$ 时隙的采样数与 $t+1-T$ 时隙采样数一致。

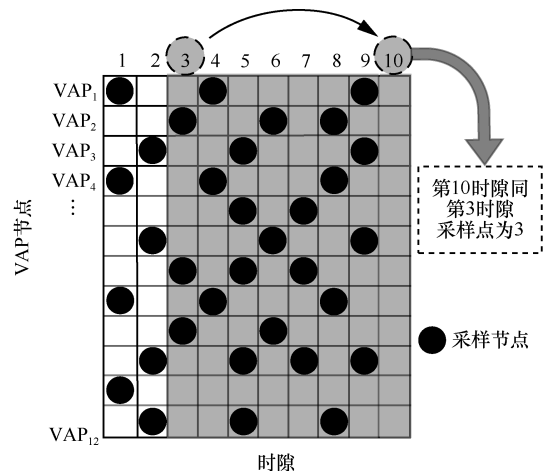


图 6 窗口固定采样

2) 自适应采样

如图 7 所示, 每一次滑动窗口内完成矩阵填充后, 将恢复的数据与上一滑动窗口恢复数据的重叠部分进行比较算出误差值, 并将之与标准误差比较, 实现采样数目的动态调整。具体算法如下。

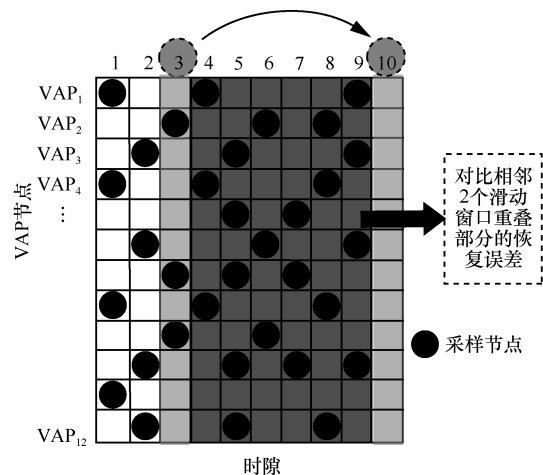


图 7 自适应连续采样

算法 1 基于随机游走的自适应连续采样算法

1) 通过窗口固定采样得到新时隙的采样点数目 m_t 。根据随机游走算法, 得到 m_t 个点的采样节点序列 D_t ;

2) 对当前窗口的数据矩阵进行矩阵填充, 得到 $\hat{X}_{N \times T}(t)$;

3) 将 $\hat{X}_{N \times T}(t)$ 与上一时隙滑动窗口填充数据 $\hat{X}_{N \times T}(t-1)$ 的交叉部分进行误差计算

$$\varepsilon_{cp} = \frac{\sum_{i=1}^N \sum_{j=t-T+1}^{t-1} |\hat{X}_{ij}(t) - \hat{X}_{ij}(t-1)|}{\sum_{i=1}^N \sum_{j=t-T+1}^{t-1} |\hat{X}_{ij}(t-1)|}$$

4) while $|\varepsilon_{cp} - \varepsilon_0| > \tau$ do

5) if $\varepsilon_{cp} - \varepsilon_0 > \tau$ then

6) 在当前时隙 t 通过随机游走算法增加 ∇ 个采样点, 此时 $m_t = m_t + \nabla$ 并更新 D_t , 重新进行矩阵填充和 ε_{cp} 误差计算;

7) else if $\varepsilon_0 - \varepsilon_{cp} > \tau$ then

8) 在下一时隙的游走采样过程中减去 ∇ 个采样点, 即 $m_{t+1} = m_{t+1} - \nabla$;

9) end if

10) 根据 m_t 和 D_t 的变化重新进行矩阵填充和 ε_{cp} 误差计算;

11) end while

12) 最终确定时隙 t 的采样节点序列 $D_t = \{x_{n+1}, x_{n+2}, x_{n+3}, \dots, x_m\}, x_i \in N$;

算法中步骤 2) 得到当前窗口的恢复矩阵 $\hat{X}_{N \times T}(t)$ 后, 将其与上一窗口的恢复矩阵 $\hat{X}_{N \times T}(t-1)$ 两者交叉的部分进行误差计算, 得到比较误差 ε_{cp} 。并与标准误差值 ε_0 比较。当 $|\varepsilon_{cp} - \varepsilon_0| > \tau$, τ 为误差阈值, 5.3 节有对其计算方法的介绍。若 $\varepsilon_{cp} - \varepsilon_0 > \tau$, 则在当前时隙通过随机游走添加 ∇ 个采样点。此时, 采样点个数增加为 $m_t = m_t + \nabla$, 同时更新采样序列 D_t , 重新进行矩阵填充并计算 ε_{cp} , 将新的比较误差 ε_{cp} 与标准误差 ε_0 进行比较, 直到两者的差距小于阈值 τ , 否则将继续添加采样点; 若 $\varepsilon_0 - \varepsilon_{cp} > \tau$, 则在下一时隙的游走采样过程中减去 ∇ 个采样点。此时, 采样点个数减少为 $m_{t+1} = m_{t+1} - \nabla$ 。根据新的采样序列进行矩阵填充并计算 ε_{cp} , 将其与标准误差进行比较, 直到两者的差距小于阈值 τ , 否则会持续更新采样序列并计算。最终时隙 t 的采样序列确定

为 $D_t = \{x_{n+1}, x_{n+2}, x_{n+3}, \dots, x_m\}, x_i \in N$ 。

5 实验方法及结果分析

5.1 测量模型

在 SDWN 框架下, 通过对传统 AP 设备的软件改造和 Libpcap、Netlink、Open vSwitch 等技术的应用, 实现了对 AP 信号强度、关联用户数、数据流量三大性能指标的数据采集。同时, 运行在应用平面的 RW-MC 算法通过 API 来处理控制器采集到的 AP 数据信息, 并指导其下达数据采集指令。最终实现全网 AP 性能数据的获取, 并记录在数据库中。如图 8 所示为测量系统的总体框架。

VAP 性能测量过程包括 2 个阶段: 学习阶段和测量阶段。学习阶段利用 RW-MC 算法确定测量样本数及样本点, 测量阶段具体为: 1) 根据学习阶段输出的测量样本率和样本点下发测量任务到数据平面; 2) 数据平面中对应的 VAP 节点执行测量任务, 并将数据返回存储在数据库中; 3) 采用奇异值阈值算法恢复测量矩阵, 恢复出的结果记录为全网 VAP 性能数据。

5.2 实验平台部署

在逸夫实验楼进行实验平台实际部署, 其真实的实验环境包括 1 个控制器和 15 个物理 AP。

图 9 中五角星符号代表实际部署的无线接入点 AP 位置, 圆点代表信号强度测试位置, 方形区域为测试的各个 AP 的信号强度 (如 309 -65 dBm, 表示在此节点测定 309 房间 AP 的信号强度为 -65 dBm)。通过 Fluke Air Check 测量的信号强度可以看出, 该区域达到了密集部署的需求。

5.3 实验测试

5.3.1 算法性能测试

1) 初始窗口矩阵采样方式对比

本实验以 AP 数据流量作为实验数据, 将改进后的模型与伯努利模型和均匀采样模型进行比较, 用 3 种方式从初始矩阵中按照采样率 28.6% 取 m 个点, 本实验中节点数目 $N=90$, 滑动窗口大小 $T=240$, 总共进行 $M=30$ 次采样。结果如图 10 所示。

由实验结果可以看出, 通过改进后的模型进行采样得出的恢复误差值不仅在整体上小于伯努利模型和均匀模型得出的结果, 且更稳定。此时, 将这 M 次恢复误差的平均值作为基准重构错误率 ε_0 , 将每一次恢复误差率与基准重构错误率的差距均值作为自适应采样的阈值 τ , 如式(16)所示。

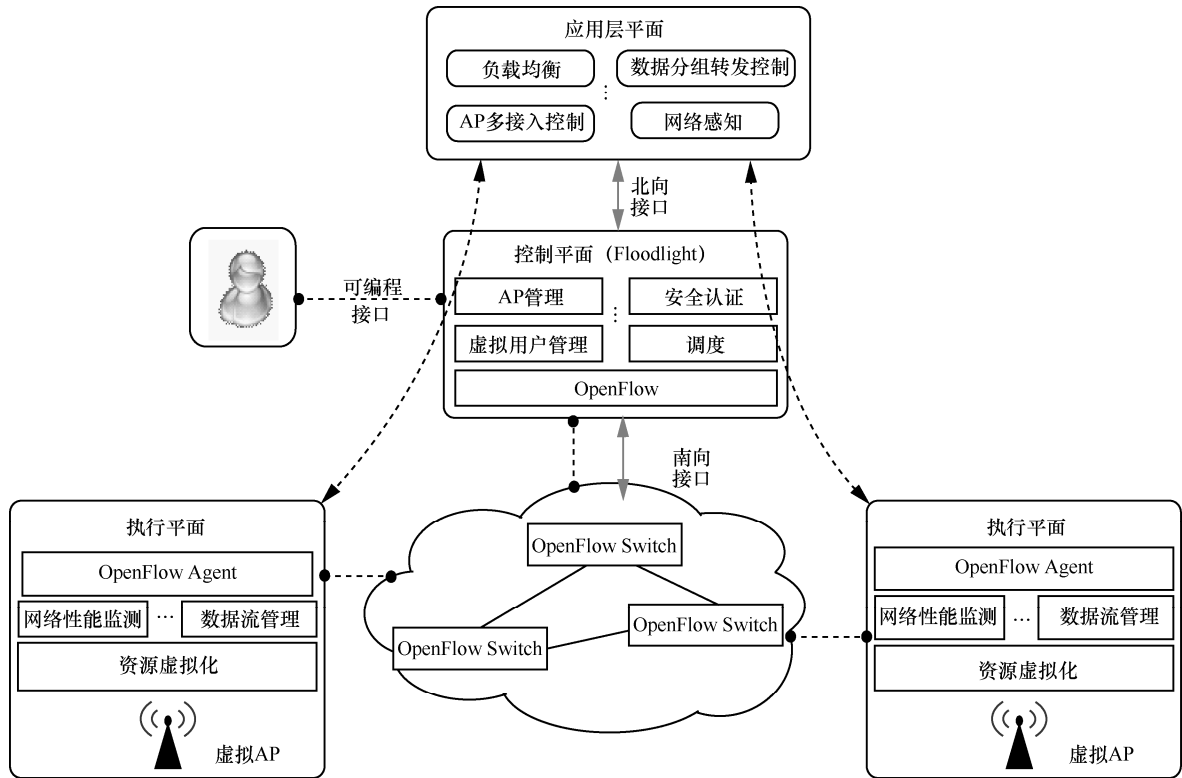


图 8 VAP 性能测量系统架构

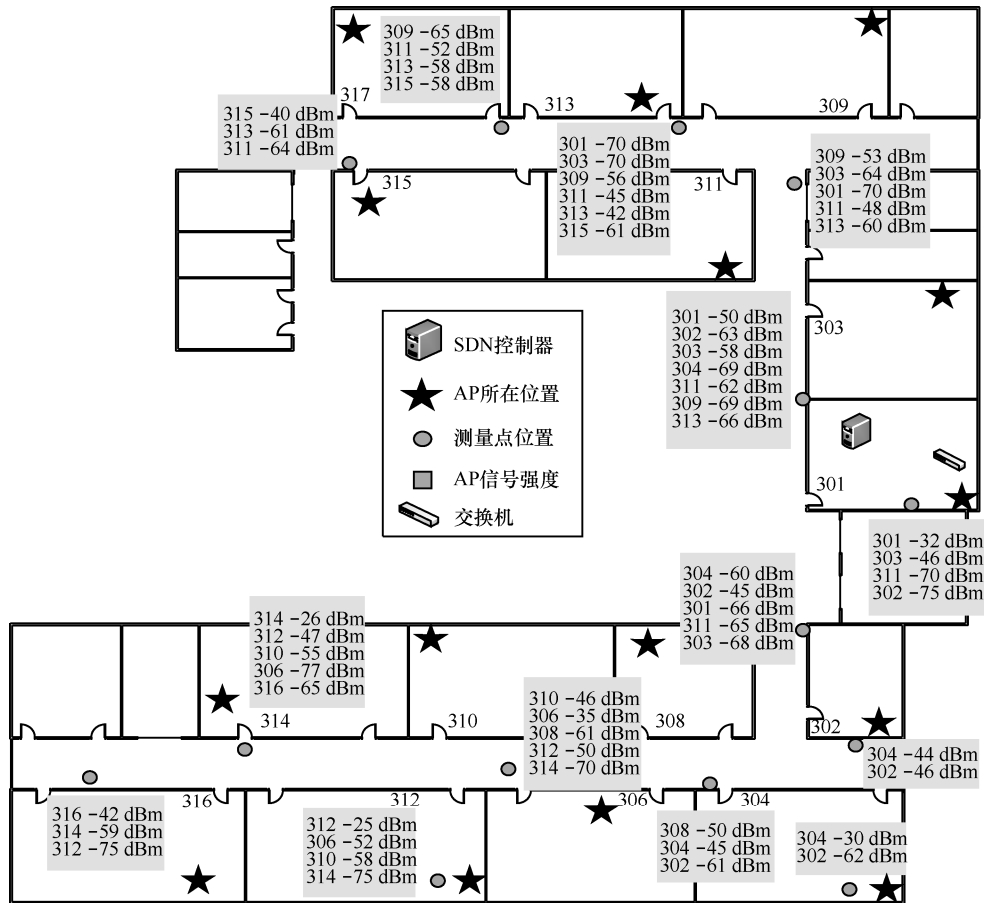


图 9 SDWN 实验平台实际部署

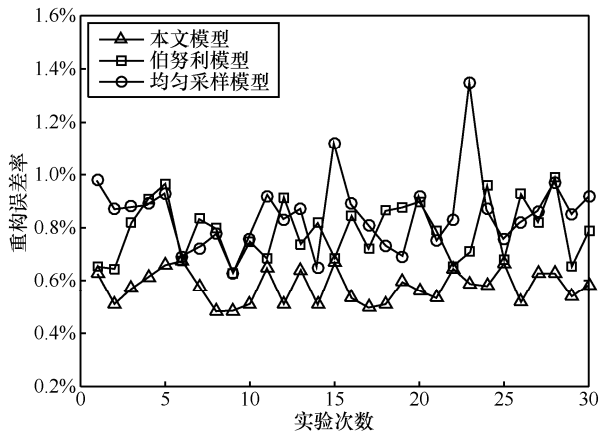


图 10 采样模型对比

$$\begin{cases} \varepsilon_0 = \frac{\sum \varepsilon}{M} \\ \tau = \frac{\sum |\varepsilon - \varepsilon_0|}{M} \end{cases} \quad (16)$$

2) 算法误差测试

本实验设置了 4 种采样方式作为对比实验，同样以 AP 数据流量作为实验数据。第一种是本文提出的基于随机游走的自适应确定采样点数；第二种是窗口固定采样数下的随机游走；第三种是窗口固定采样数下的交叉采样，交叉采样即相邻的 2 个时隙采样节点不重复；第四种是每一时隙按照 20%、30%、40% 的固定比例采样，采样方式为随机采样。实验中初始矩阵的采样率在 28% 左右。最终将以上 4 种方式得到的重构数据矩阵与整体测量数据矩阵比较，得出的重构误差如图 11 所示。

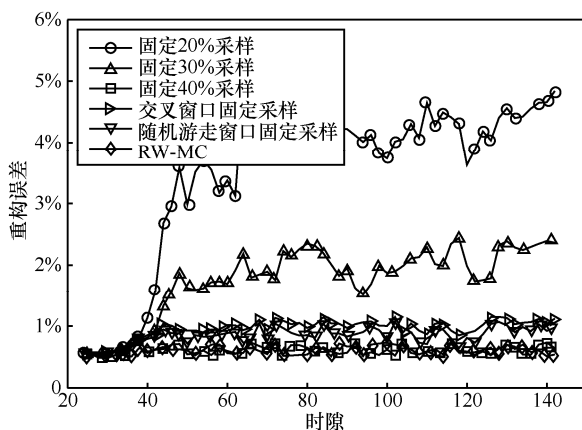


图 11 恢复误差对比

本文自适应采样算法的重构误差为 0.6%~0.7%，与固定采样率 40% 的效果较为接近，重构误差远远小于固定采样率为 20% 和 30% 的情况，证明了本文

算法低采样、高准确的特点；同时，将固定窗口采样数与随机游走和交叉采样 2 种方式对比可以看出，随机游走方式能结合历史信息找出最优测量点，实现更高的恢复准确率；最后，自适应算法比随机游走窗口固定采样方式的重构误差小 0.3% 左右，证明了自适应算法的有效性。

5.3.2 系统性能测试

1) 流量对比实验

图 12 为实验平台内中心交换节点的数据流量随时间的变化情况，系统内 AP 均通过该节点实现与控制器和互联网的数据交换。实验测试过程中，为每个 VAP 施加 10 Mbit/s 的数据流作为实验的背景流量。通过在相同测试环境下对 VAP 的网络性能采用全网方法和 RW-MC 方法进行实际测量，对比 2 种方法的测量结果，可以看出全网测量消耗的流量是 RW-MC 测量方式的 3 倍。

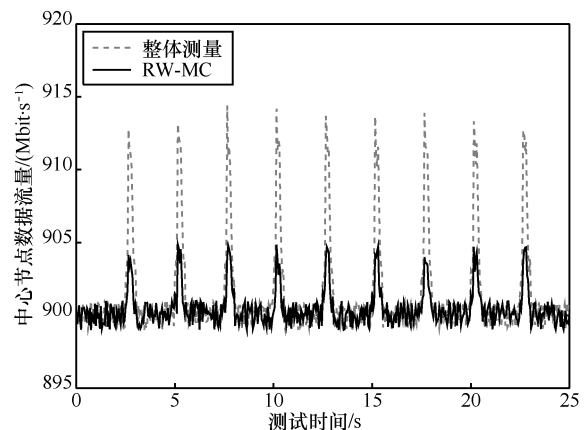


图 12 中心节点数据流量对比

6 结束语

针对全网 VAP 性能测量产生的存储资源不足、处理能力不够、测量数据传输占用大量带宽等问题，提出了一种基于随机游走的矩阵填充算法。在保证低重构误差的情况下，减少直接测量样本数，降低测量数据占用的存储空间大小和降低测量数据传输占用的带宽。同时，测量结果有利于研究者解决负载均衡、切换和节能等网络优化问题。

参考文献:

[1] RIGGIO R, MARINA M K, SCHULZ-ZANDER J, et al. Programming abstractions for software-defined wireless networks[J]. IEEE Transactions on Network and Service Management, 2015, 12(2): 146-162.

- [2] NGUYEN T T, BONNET C, HARRI J. SDN-based distributed mobility management for 5G networks[C]//Wireless Communications and Networking Conference (WCNC).2016: 1-7.
- [3] AKYILDIZ I F, NIE S, LIN S C, et al. 5G roadmap: 10 key enabling technologies[J]. Computer Networks, 2016, 106: 17-48.
- [4] SCHULZ-ZANDER J, MAYER C, CIOBOTARU B, et al. OpenSDWN: programmatic control over home and enterprise Wi-Fi[C]//Proceedings of the 1st ACM SIGCOMM Symposium on Software Defined Networking Research. 2015: 16.
- [5] TANAKA T, ABE K, AUST S, et al. Automatic and cooperative sleep control strategies for power-saving in radio-on-demand WLANs[C]//Green Technologies Conference. 2013: 293-300.
- [6] LIU Z, LIU Y, GONG Z, et al. A muti-rate access point selection policy in IEEE 802.11 WLANs[C]// 2011 International Conference on Multimedia Technology (ICMT). 2011: 63-67.
- [7] MO Y, CAI Y, WANG B. A novel indoor localization method based on virtual AP estimation[C]//IEEE International Conference on Communications. 2012: 5508-5512.
- [8] HAN Z, LI H, YIN W. Compressive sensing for wireless networks[M]. Cambridge University Press, 2013.
- [9] QI X, WANG Y, WANG Y, et al. Compressive sensing over strongly connected digraph and its application in traffic monitoring[C]// IEEE INFOCOM 2014 - IEEE Conference on Computer Communications. 2014: 446-453.
- [10] MALBOUBI M, WANG L, CHUAH C N, et al. Intelligent SDN based traffic (de) aggregation and measurement paradigm (iSTAMP)[C]// IEEE INFOCOM 2014 - IEEE Conference on Computer Communications. 2014: 934-942.
- [11] KESHAVAN R H, OH S, MONTANARI A. Matrix completion from a few entries[C]//2009 IEEE International Symposium on Information Theory. 2009: 324-328.
- [12] MALBOUBI M, GONG Y, XIONG W, et al. Software defined network inference with passive/active evolutionary-optimal probing (SNIPER)[C]// International Conference on Computer Communication and Networks. 2015: 1-8.
- [13] XIE K, WANG L, WANG X, et al. Sequential and adaptive sampling for matrix completion in network monitoring systems[C]//2015 IEEE Conference on Computer Communications (INFOCOM). 2015: 2443-2451.
- [14] HE J, SUN G, ZHANG Y, et al. Data recovery in wireless sensor networks with joint matrix completion and sparsity constraints[J]. IEEE Communications Letters, 2015, 19(12): 2230-2233.
- [15] CHENG J, YE Q, JIANG H, et al. STCDG: an efficient data gathering algorithm based on matrix completion for wireless sensor networks[J]. IEEE Transactions on Wireless Communications, 2013, 12(2): 850-861.
- [16] XIE K, WANG L, WANG X, et al. Learning from the past: intelligent on-line weather monitoring based on matrix completion[C]//2014 IEEE 34th International Conference Distributed Computing Systems (ICDCS). 2014: 176-185.
- [17] RECHT B, FAZEL M, PARRILO P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization[J]. SIAM Review, 2010, 52(3): 471-501.
- [18] FENG W, JING C, LI L. Recommender system based on random walk with topic model[C]//International Conference on Advanced

Computing. 2016.

- [19] SHI R, GOSWAMI M, GAO J, et al. Is random walk truly memoryless- traffic analysis and source location privacy under random walks[J]. Proceedings - IEEE INFOCOM, 2013, 12(11): 3021-3029.
- [20] CANDÈS E J, TAO T. The power of convex relaxation: near-optimal matrix completion[J]. IEEE Transactions on Information Theory, 2010, 56(5): 2053-2080.

作者简介:



王新恒(1988-), 男, 山东枣庄人, 重庆邮电大学博士生, 主要研究方向为网络测量、SDN、NFV 以及无线通信。



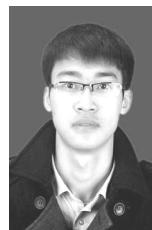
王倩云(1992-), 女, 四川渠县人, 重庆邮电大学硕士生, 主要研究方向为无线局域网节能、网络测量。



王佳杰(1992-), 男, 河北唐山人, 重庆邮电大学硕士生, 主要研究方向为网络测量、无线通信。



赵国锋(1972-), 男, 陕西西安人, 博士, 重庆邮电大学教授, 主要研究方向为未来网络、SDN、NFV、网络安全等。



靳文强(1993-), 男, 内蒙古乌兰察布人, 重庆邮电大学硕士生, 主要研究方向为未来网络、SDN、NFV 以及无线通信。