

业务驱动的密集移动终端网络资源预留方法

何 芬^a, 程良伦^b, 王 涛^a

(广东工业大学 a. 自动化学院; b. 计算机学院, 广州 510006)

摘 要: 为提高密集移动终端商圈区域的终端服务质量, 提出一种基于业务驱动资源预留的密集移动终端网络接入方法。根据网络属性、终端运动预测轨迹和业务类型, 为高优先级业务终端预留资源量, 在尽可能少牺牲低优先级业务服务质量的前提下, 使高优先级业务终端获得稳定可靠的网络服务, 并通过时间触发公平机制保障低优先级业务终端的网络接入资源。仿真结果表明, 该方法不仅能够保障高优先级业务移动终端在异构网络中的接入质量, 提高网络资源利用率, 而且可以降低热点高容量场景下低优先级业务终端的接入失败率, 改善用户体验质量。

关键词: 密集移动终端; 异构网络接入; 业务驱动; 运动轨迹预测; 资源预留

中文引用格式: 何芬, 程良伦, 王涛. 业务驱动的密集移动终端网络资源预留方法[J]. 计算机工程, 2019, 45(9): 87-94.

英文引用格式: HE Fen, CHENG Lianglun, WANG Tao. Business-driven resource reservation method for dense mobile terminal network[J]. Computer Engineering, 2019, 45(9): 87-94.

Business-driven Resource Reservation Method for Dense Mobile Terminal Network

HE Fen^a, CHENG Lianglun^b, WANG Tao^a

(a. School of Automation; b. School of Computers, Guangdong University of Technology, Guangzhou 510006, China)

[Abstract] In order to improve the Quality of Service (QoS) of terminal in the area of dense mobile terminal business district, a network access method for dense mobile terminal network based on business-driven resource reservation is proposed. It reserves resources for terminals with high-priority business based on the attributes of networks, the forecasted motion trajectory of terminals and the types of business. Under the premise of minimizing the sacrifice of QoS of the low-priority service, it ensures high-priority business terminals get a smooth and reliable network service, and ensures network access resources for terminals with low-priority business by using time-triggered fairness mechanism. Simulation results show that, this method can not only ensure the access quality of mobile terminals with high-priority business in heterogeneous network and improve the utilization of network resources, but also reduce the access failure rate of low-priority business terminals in hot-spot high-capacity scenarios. What's more, it can improve the users' Quality of Experience (QoE).

[Key words] dense mobile terminal; heterogeneous network access; business-driven; motion trajectory forecast; resource reservation

DOI: 10.19678/j.issn.1000-3428.0052263

0 概述

随着 5G 网络的发展, 移动终端数量爆发式增长, 终端密集的热点高容量地区网络资源竞争激烈^[1-2], 如大型商圈区域同时存在上万个移动终端需要接入网络, 其中可靠性、实时性要求高的关键性业务, 需要平滑、稳定的网络服务质量 (Quality of Service, QoS)^[3], 而移动终端频繁的网络切换会导致较大的资源损耗和网络接入时延, 难以保障终端的网络服务质量^[4-5]。目前, 诸多网络接入方法被提出以提高终

端服务质量, 如基于优先级网络资源抢占^[6]、最优化网络资源分配^[7]和网络资源预留^[8]等, 其中网络资源预留方法根据预定规则为特定的终端预留网络资源, 能够保证指定终端的网络接入和服务质量。

文献[9]提出实时按照业务类型的变化改变资源预留的方案, 通过路由和频谱分配解决因资源预留请求导致的即时预留服务降级问题, 以实现动态资源分配的, 但该方案未考虑网络服务质量的优化。文献[10-12]通过建立 QoS 与预留资源之间的关系, 提出负载均衡的多资源节点协同预留方案。文

基金项目: 国家自然科学基金 (61672170); 国家自然科学基金青年科学基金 (61502110); 广东省重大科技专项 (2015B010104005); NSFC-广东联合基金 (U1701262); 广东省省级科技计划 (2017A010101017)。

作者简介: 何 芬 (1993—), 女, 硕士研究生, 主研方向为异构网络数据融合; 程良伦, 教授、博士生导师; 王 涛, 副教授、博士研究生。

收稿日期: 2018-07-30 **修回日期:** 2018-09-10 **E-mail:** 2298103602@qq.com

献[13-14]引入自适应门限值对网络资源进行自适应调整,以提高网络资源利用率并避免流量匮乏的问题,但不能使用户体验质量(Quality of Experience, QoE)得到明显改善。文献[15]提出基于排队机制的动态资源预留算法,优先保障切换终端的接入,对于新用户则是概率接入,当系统资源耗尽时采取排队机制,等待系统有资源时再进行接入。文献[16]动态更新网络带宽和计算资源,在线修订资源预留方案,但该方法存在延迟较大的问题,难以及时调整资源预留方案。文献[17]使用大数据分析预测终端运动行为,根据预测结果进行资源预留,减少了传输时间。文献[18]利用业务特点提出软资源预留机制以减少延迟,实现运营商与远程设备之间的通信,但该方法仅适用于小数据量资源预留。文献[19]通过云服务提供商最优化带宽预留方案,设计预测性资源自动扩展系统,保证流媒体性能,但在资源紧缺的地区,将大量资源用于保障流媒体性能会降低其他业务的服务质量。

上述研究中的资源预留方法或设置固定资源预留门限值,或通过预测终端数量决定网络资源预留量,使资源预留受场景的影响较大,无法适应突发情况。当资源紧缺时,切换终端或新终端只能等待其他终端释放资源,降低了用户体验质量。针对该问题,本文设计一种密集移动终端下业务驱动的网络资源预留方法。该方法预测终端运动轨迹和关键性终端的网络持续时间,并基于用户偏好为关键性终端预选择目标网络,以提前预留资源。在此基础上,根据时间触发公平性原则均衡各优先级终端网络接入量,避免终端出现“饿死”现象。

1 业务驱动的资源预留机制

在如大型商圈的密集高容量区域,终端密度、位置、运行的业务类型随时间和空间的变化均可能发生变化,因此,终端对网络资源的需求也随之动态变化^[20]。在商圈区域运行的业务类型一般有实时支付、在线会话、网络视频和其他后台业务等,本文将运行实时性要求高的实时支付、在线会话等业务的终端归类为高优先级终端,反之则为低优先级终端。本文方法为高优先级终端提前预留资源,其他正运行低优先级业务的终端依然保留资源竞争的方式进行网络接入。

业务驱动资源预留机制的整体框架如图1所示。在移动终端运动模块中根据高优先级终端的初速度、加速度和运动方向实时预测终端的运动轨迹,计算符合终端需求的备选网络,并通过目标网络预测模块确定为终端预留资源的目标网络。当终端离开原网络的覆盖范围、原网络无法提供满足要求的网络资源或终端运行的业务由高优先级转为低优先级时,释放原预留的网络资源。若所有的网络资源均优先满足高优先级终端,则低优先级终端容易出现“饿死”现象,造成严重的网络服务不公平性。因此,当网络中低优先级终端使用的网络资源占该网络总资源率低于设定的阈值 Th 时,利用队列记录被抢占网络资源的低优先级终端。队列中的低优先级终端可优先使用网络中其他终端完成数据传输后释放出的网络资源,时间触发队列中的低优先级终端获得网络服务资源,可提高低优先级终端的接入率。

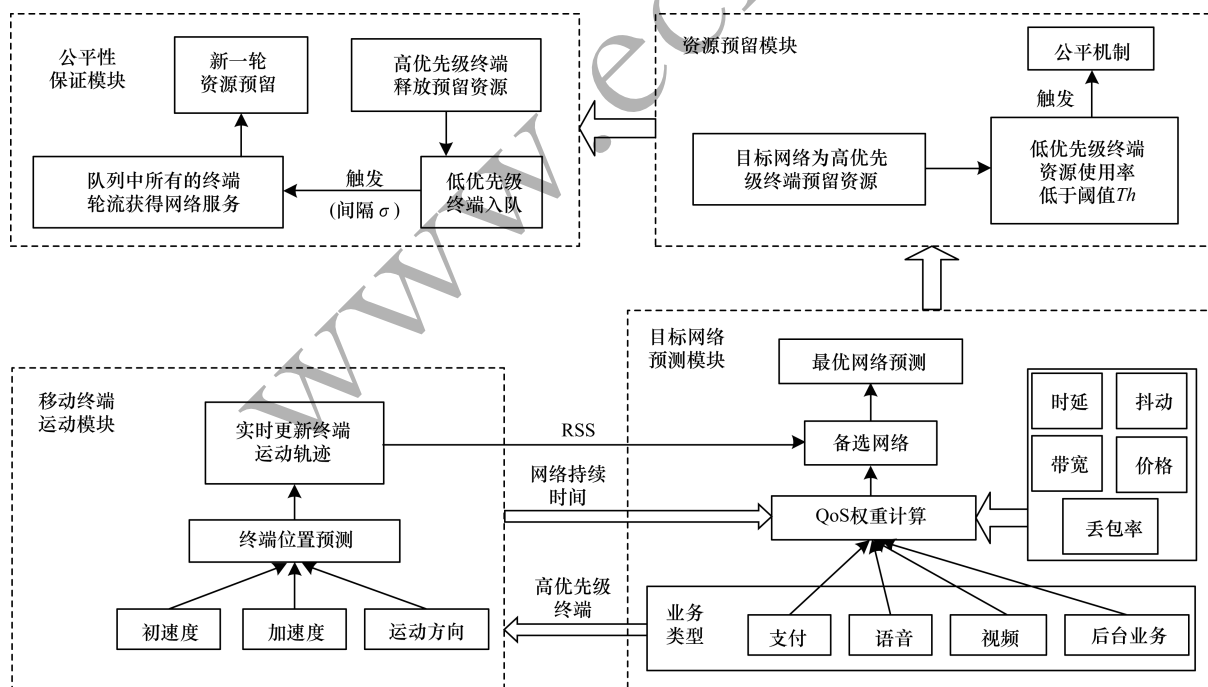


图1 业务驱动资源预留机制的整体框架

2 移动终端运动轨迹预测

终端在商圈区域运动状态随机性较大,为模拟终端的随机运动,必须综合考虑终端当前的运动速度、加速度(改变终端当前运动速度和方向),分析不同运动状态终端在各网络中持续时间,明确运动状态对网络选择的影响。

2.1 终端运动模型

在异构网络环境下选择接入网络时,本文通过建立一个终端运动模型进行分析,将终端在短时间内的运动轨迹看作一条直线,终端的运动趋势和接入点的关系如图2所示,其中, D_1 、 D_2 分别表示终端与基站A、B之间的距离。当终端从位置1移动到位置2时, D_1 逐渐增大。如果2个网络性能相同,终端优先选择接入到即将接近的网络,以增长网络持续时间,避免乒乓效应。

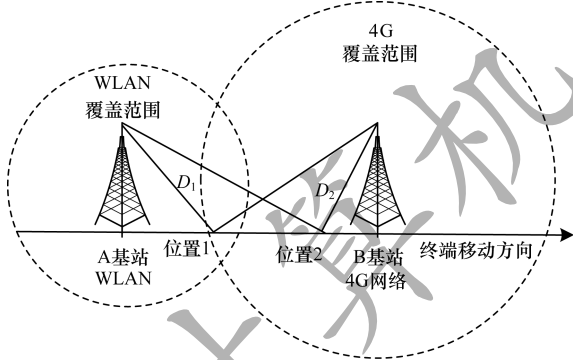


图2 移动终端运动模型

2.2 移动终端轨迹及持续时间预测

终端运动轨迹预测法通过获取终端接收信号强度(Received Signal Strength, RSS)值,计算当前终端初始速度。当RSS超过预设阈值时,该网络成为移动终端预选选择候选网络之一。利用终端自带的加速度仪和陀螺仪测量加速度辅助速度的测量,实时估计终端的运动趋势,有效减少计算量。终端RSS值的计算公式为^[21]:

$$R(d) = K_1 - K_2 \lg(d) + u(x) \quad (1)$$

其中, K_1 、 K_2 分别是网络发送功率和路径损失因子, d 是终端与基站之间的距离, $u(x)$ 为服从参数 $(0, \sigma_1)$ 的高斯随机分布函数。

若高优先级终端与当前所接入网络基站之间的距离 d_{switch} 接近该网络的覆盖半径,则对该高优先级终端进行运动轨迹预测及网络持续时间预测。

抽样时间 T_d 内距离的累积变化和为:

$$\Delta D_d = 10^{\frac{K_1 + u(x) - R_{vi} + 1}{K_2}} - 10^{\frac{K_1 + u(x) - R_v}{K_2}} \quad (2)$$

终端在采样时间内的初始速度为:

$$v_1 = \Delta D_d / t_{\text{inter}} \quad (3)$$

其中, v_1 为终端的初始计算所得的速度, t_{inter} 为相邻2个抽样的时间间隔。

利用 t 时刻检测到的终端实时加速度 a_t 预估终端实时速度,当加速度方向与当前速度方向不同时,终端速度大小及方向发生改变,终端运动轨迹随之改变,如式(4)所示。

$$v_t = \sqrt{(v_{t-1} + a_t \cos \theta_{t_{\text{inter}}})^2 + (a_t \sin \theta_{t_{\text{inter}}})^2} \quad (4)$$

其中, $t \geq 2$, v_t 、 v_{t-1} 分别为 t 时刻预估的终端实时速度和前一抽样间隔预估的终端实时速度, θ 为加速度 a_t 与速度 v_{t-1} 之间夹角。当检测到的加速度 a_t 的值大于阈值 $a_{\text{threshold}}$ 或夹角 θ 大于阈值 $\theta_{\text{threshold}}$ 时(终端运动轨迹发生明显变化),再次预测终端运动轨迹。

根据终端与所接入基站之间距离为 d_{switch} 时所在的位置和实时速度 v_t 预测当前终端的运动轨迹。以基站为原点建立一个网络持续时间的预测模型,如图3所示。

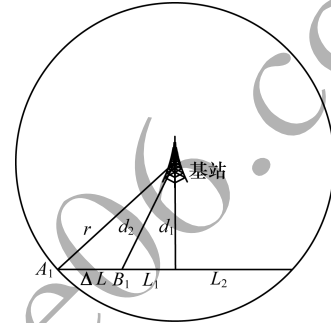


图3 网络持续时间预测模型

假设 A_1 为预测轨迹与网络边缘交点, B_1 为预测轨迹上的任意一点, r 为该网络覆盖半径, d_1 为接入点与预测轨迹之间的最短距离, d_2 为接入点与 B_1 之间的距离,则网络的持续时间为:

$$\begin{cases} r^2 = d_1^2 + (\Delta L + L_1)^2 \\ d_2^2 = d_1^2 + L_1^2 \end{cases} \quad (5)$$

由此可得:

$$L_1 = (r^2 - d_2^2 - \Delta L^2) / (2\Delta L), \Delta L = \int_0^{\Delta t} v_t dt$$

因为 $L_2 = L_1 + \Delta L$,所以预测网络的持续时间为:

$$t = \frac{2L_2}{V_{\text{total}}} = \frac{[r^2 - d_2^2 + (\int_0^{\Delta t} v_t dt)^2]}{\int_0^{\Delta t} v_t dt / \Delta t} \quad (6)$$

其中, Δt 为任意2个时间间隔, V_{total} 为该段时间内的平均速度。根据式(6)计算出的预估网络持续时间,实时监测终端的运动速度。

3 业务驱动的最优网络选择

本文通过终端轨迹预测确定符合终端要求的候

选网络集合,由于终端用户对目标网络要求不同,如网络持续时间、带宽或费用优先等,因此会造成最终选择的目标网络存在一定的差异。根据终端用户的偏好,基于终端当前运行的业务类型,利用层次分析法确定基于网络判决属性参数的权重和多业务类型的优先级权重,采用逼近理想值法对所有候选网络加权排序,以此选择出最优接入网络作为目标网络,并在目标网络中按照高优先级终端的资源需求预留网络资源。

3.1 基于用户偏好的网络属性权重确定

本文主要考虑的 6 种网络属性分别为时延、抖动、带宽、丢包率、价格和网络持续时间,其中高优先级终端偏好于低时延、低抖动和低丢包率的网络。为确定各属性参数对高优先级终端选择最优网络的影响程度,对上述 6 种网络属性采用九级标度法构造高优先级终端业务判断矩阵 $R^{[22]}$:

$$R = \begin{bmatrix} 1 & 3 & 5 & 3 & 5 & 2 \\ 1/3 & 1 & 3 & 1 & 3 & 1/2 \\ 1/5 & 1/3 & 1 & 1/3 & 1 & 1/4 \\ 1/3 & 1 & 3 & 1 & 3 & 1/2 \\ 1/5 & 1/3 & 1 & 1/3 & 1 & 1/4 \\ 1/2 & 2 & 4 & 2 & 4 & 1 \end{bmatrix} \quad (7)$$

令 R_{ij} 是矩阵 R 中的元素值,为每两个属性元素之间的相对重要程度之比。设 $Rw = \lambda_{\max} w$,得:

$$w_i = \prod_{j=1}^n R_{ij}, i=1,2,\dots,n \quad (8)$$

其中, λ_{\max} 为矩阵 R 的最大特征根, n 为属性元素的个数, w 为对应的特征向量。

对权重进行归一化处理,得到高优先级终端 QoS 判决属性参数权重向量 U ,其中元素 u_i 计算公式如下:

$$u_i = \frac{w_i}{\sum_{i=1}^n \left(\sqrt[n]{\prod_{j=1}^n w_j} \right)} \quad (9)$$

则 $U = [0.370 \quad 0.141 \quad 0.056 \quad 0.141 \quad 0.056 \quad 0.236]^T$ 。

3.2 最优网络选择

由当前网络环境下监测到的候选网络参数值可得多属性决策矩阵为 $C = (c_{ij})_{p \times q}$,其中, p 为候选网络数量, q 为判决元素数量。

1) 对多属性决策矩阵进行规范化处理。

对于效益型指标(如带宽),需要其值越大越好,即:

$$k_{ij} = \frac{c_{ij}}{\max\{c_{ij} | 1 \leq i \leq p\} + \min\{c_{ij} | 1 \leq i \leq p\}} \quad (10)$$

对于成本型指标(如时延、价格等),则需要其值越小越好,即:

$$k_{ij} = \frac{\max\{c_{ij} | 1 \leq i \leq p\} + \min\{c_{ij} | 1 \leq i \leq p\} - c_{ij}}{\max\{c_{ij} | 1 \leq i \leq p\} + \min\{c_{ij} | 1 \leq i \leq p\}} \quad (11)$$

2) 规范化判决矩阵为 $Z = (z_{ij})_{p \times q}$,有:

$$z_{ij} = \frac{k_{ij}}{\sum_{i=1}^p k_{ij}} u_i, i=1,2,\dots,p, j=1,2,\dots,q \quad (12)$$

3) 对所有候选网络按接近理想值的程度进行排序,选取最接近理想值的网络,即选择 C_{T_i} 值最大的网络:

$$C_{T_i} = \frac{S_i^+}{S_i^+ + S_i^-} \quad (13)$$

其中, S_i^+ 和 S_i^- 分别是正理想解和负理想解,计算公式如式(14)和式(15)所示。

$$S_i^+ = \sqrt{\sum_{j=1}^q (z_{ij} - z_j^+)^2} \quad (14)$$

$$S_i^- = \sqrt{\sum_{j=1}^q (z_{ij} - z_j^-)^2} \quad (15)$$

在式(14)和式(15)中, z_j^+ 为参数 j 的最佳值。

当 z_{ij} 为收益型参数时, z_j^+ 取最大值;当 z_{ij} 为开销型参数时, z_j^+ 取最小值,而 z_j^- 为参数 j 的最差值。

4 网络资源预留方法

4.1 目标网络资源预留

根据网络状态和终端用户偏好确定目标网络后,目标网络为该高优先级终端预留网络资源。网络中的资源可分为 3 个部分:

1) 为防止网络资源使用过饱和,每个网络保留一部分资源 $S_{\text{saturation}}$ 。当网络中剩余资源少于 $S_{\text{saturation}}$ 时,则仅允许高优先级切换终端接入网络,保障高优先级终端获得稳定、平滑的网络服务。

2) 基于资源预留机制,高优先级业务使用的资源 S_{reserve} ,即本文算法需要调整的网络资源。

3) 通过资源预留后网络中剩余的资源 S_{low} ,所有终端业务均可以资源抢占方式按需选择接入。

本文算法中调整预留网络资源部分具体表现为:当高优先级终端 1 进入多个网络的覆盖时,基于移动终端运动轨迹预测和业务驱动的最优网络选择法选择出目标网络 A,根据运行的业务类型确定终端 1 所需的网络资源量,此时网络 A 为该终端提前预留符合需求的网络资源,其他正在等待接入网络的低优先级终端暂时无法使用预留的网络资源。而商圈区域内存在多个不同的网络,且每个网络需要为多个终端预留资源。因此,假设商圈中网络有 M 个,待接入的高优先级终端总数为 N ,每一网络预接入的高优先级终端数为 n_1 ,低优先级终端数为 n_2 ,依然保留资源竞争方式进行网络接入。设 S_{bw_j} 为网络 i 给终端 j 预留的网络资源,则第 i 个网络预留资源量为:

$$S_{\text{reserve}_i} = \sum_{j=1}^{n_1} S_{bw_j}$$

$$\text{s.t. } 0 \leq S_{\text{reserve}_i} \leq S_{\text{total}_i}, \forall i = 1, 2, \dots, M \quad (16)$$

$$\sum_{i=1}^M (n_1 + n_2) = N$$

$$0 \leq \sum_{i=1}^M n_1 \leq N, 0 \leq \sum_{i=1}^M n_2 \leq N \quad (17)$$

其中, S_{total_i} 、 S_{reserve_i} 分别代表网络 i 的总资源量和为高优先级终端预留的资源量。

然而移动终端的实际运动轨迹和运行的业务类型具有一定的随机性,例如,当终端 1 的运动方向或者运行的业务类型突然发生改变,会造成原先预留的网络资源浪费,则在整个商圈区域内,所有网络预留资源的总浪费浪费率为 R_{waste} :

$$R_{\text{waste}} = \sum_{i=1}^M (S_{\text{reserve}_i} - \sum_{j=1}^{n_{\text{exact}}} S_{ij}) / S_{\text{total}_i} \quad (18)$$

其中, n_{exact} 为最终进入预选择网络的高优先级终端数, S_{ij} 是高优先级终端 j 实际占用网络 i 的资源量。为解决上述的资源浪费问题,网络为移动终端预留资源时加上时间戳 T_{stamp} , 当终端在预测的时间内没有到达目标网络,即当检测到终端运动轨迹或运行的业务发生改变,则目标网络释放为该终端预留的资源。

4.2 时间触发的公平性保障

若通过资源预留机制将所有的网络资源均预留给高优先级终端,当高优先级终端数过多时,则可能造成低优先级终端处于持续的等待阶段,甚至出现“饿死”现象。

按照实际的业务服务分析,高优先级的支付业务和语音业务的数据传输量较少,服务时间短;而低优先级的视频业务数据传输量大,且服务时间长。若在资源紧缺的情况下,等待接入网络的低优先级终端完成数据传输,不但不能改善低优先级终端的用户体验质量,而且还会降低高优先级终端的接入率。因此,本文提出时间触发的公平性保障机制,分以下 2 种情况进行处理:

1) 当网络中所有低优先级终端占用的网络资源率高于阈值 Th 时,高优先级终端确定目标网络后,抢占网络中低优先级终端网络资源,为高优先级终端预留足够的网络资源量,保障高优先级终端网络服务,而被抢占网络资源的低优先级终端则需要等待网络中有空闲资源后再通过传统的资源抢占方法获得网络服务资源。

2) 当网络中所有低优先级终端占用的网络资源率低于阈值 Th 时,该网络触发公平性机制,将触发时刻标记为 T ,并依次用队列记录所有被抢占网络资

源的低优先级终端,等待网络中其他终端完成数据传输释放出网络资源后,按照队列的先进先出原则,队列中最先记录的低优先级终端获得网络接入优先权,优先获得网络服务,即队列中最先入队的低优先级终端 u_{low_1} 获得网络资源。以时间间隔 δ 作为每个终端获得网络服务的时长,在 $T_1 = T + \delta$ 时刻后, u_{low_1} 失去网络使用权,队列中下一个终端随之等待接入网络。以此类推,直至队列中所有低优先级终端均获得网络服务机会后,进入新一轮的资源预留。当网络中低优先级终端的网络资源使用率超过 Th 后,取消时间触发公平性机制。

5 仿真实验

本节通过模拟移动终端密集的商圈来验证所提出的方法。仿真实验分别针对不同网络资源预留量、预留资源浪费量、不同优先级业务接入失败率和网络响应率等进行验证,通过与传统的资源抢占机制和固定预留资源门限值算法对比,评估本文方法的性能。

5.1 实验场景

实验中模拟一个 $3.5 \text{ km} \times 3.5 \text{ km}$ 密集高容量大型商圈区域内的异构网络环境,包括 4G 网络和无线局域网 (Wireless Local Area Networks, WLAN) 2 种制式的网络类型,如图 4 所示,其中,星点代表商圈中终端用户位置,呈高斯分布,越靠近商场中心区域终端越密集。

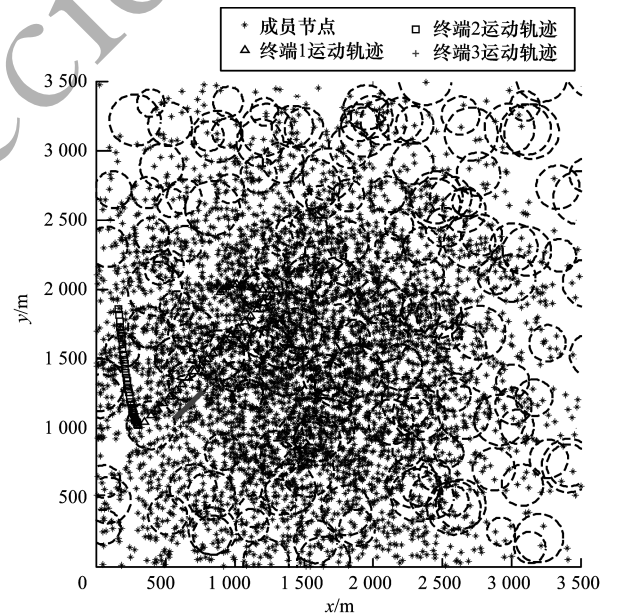


图 4 实验场景

本文主要考虑终端用户有 4 种运动状态:高速 ($6 \text{ m/s} \sim 7 \text{ m/s}$, 终端 1), 中速 ($3 \text{ m/s} \sim 4 \text{ m/s}$, 终端 2), 慢速 ($1 \text{ m/s} \sim 2 \text{ m/s}$, 终端 3), 静止 (0 m/s), 并

为终端设置不同的加速度值,实现其加减速和改变运动方向。图 4 中虚线圆圈代表每个 WLAN 网络的覆盖范围,WLAN 网络的分布同样服从高斯分布,4G 网络为商圈内全覆盖。具体的参数设置如表 1 所示,其中 r 代表 WLAN 网络的覆盖半径,

各网络参数随着覆盖范围的变化而变化。根据实际的商圈环境,运行低优先级业务比高优先级业务的终端多,其中支付业务需要的带宽为 15 KB/s,语音业务为 60 KB/s,视频业务为 100 KB/s,后台业务为 30 KB/s。

表 1 网络参数设置

网络属性	时延/s	抖动/s	带宽/(KB·s ⁻¹)	丢包率/(bit·s ⁻¹)	价格/元
4G 网络	0.20	0.020	100 000	1 500.0	1.0
WLAN(100 < r < 120)	0.05 ~ 0.06	0.005 ~ 0.045	500	5.0	0.1
WLAN(120 < r < 140)	0.06 ~ 0.07	0.005 ~ 0.045	1 500	15.0	0.2
WLAN(140 < r < 160)	0.07 ~ 0.08	0.005 ~ 0.045	2 500	25.0	0.3
WLAN(160 < r < 200)	0.08 ~ 0.10	0.005 ~ 0.045	62 500	62.5	0.5

5.2 仿真结果分析

假设当前在商圈中的移动终端为 8 000 个,每个网络根据终端分布情况及移动轨迹为高优先级的终端预留的资源量如图 5 所示。实验中共模拟了 178 个不同网络,图 5 中仅展示了其中 10 个网络的资源预留情况。由于支付及语音业务更为倾向于短时限、持续时间长及少掉话率,因此编号为 178 的 4G 网络预留的资源较多,预留资源占总资源的 40% 左右。而低优先级的视频及后台业务时延要求低且需要的带宽较大,价格相对于 4G 网络低廉的 WLAN 更为适合。

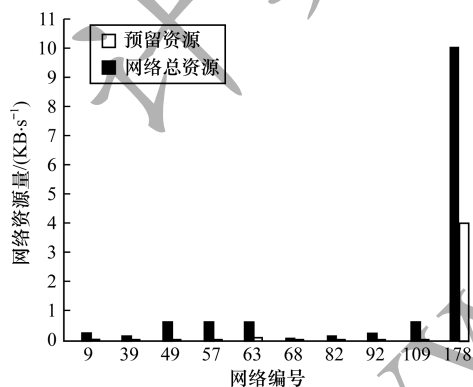


图 5 预留资源量与网络总资源量比较

通过改变实验区域内包含的终端数量来检验资源预留机制的性能。如图 6 所示,随着终端用户的位置变化,终端选择的网络随之变化,网络为高优先级终端预留的资源量会做出相应的调整。初始阶段进入实验区域的终端数量不同可以模拟出商圈内淡旺季场景,从图 6 可以看出,网络资源预留机制会根据高优先级终端数量调整网络资源预留量。当仅有 2 000 个终端时,4G 网络仅预留 10% 左右的资源给高优先级用户,而当终端数达到 10 000 个时,资源预留量达到 50% 左右。

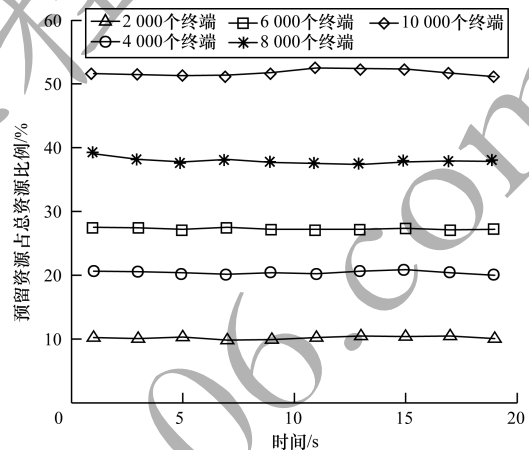


图 6 预留资源调整示意图

本文方法在不同资源预留时的资源浪费率比较如图 7 所示。每一固定资源预留量都有一个最适合的终端密集度,如终端数量为 2 000 ~ 4 000 时,预留 10% 的资源可使预留资源浪费率达到最低,而当终端数量超过 4 000 时,出现部分高优先级终端需要通过资源竞争的方式接入网络,降低了高优先级终端的接入率。而商圈地区在不同时间段内终端密集程度不同,本文提出的方法可以根据终端密集度作出资源预留率的弹性改变,在减少资源浪费率的同时保证高优先级终端的接入率。

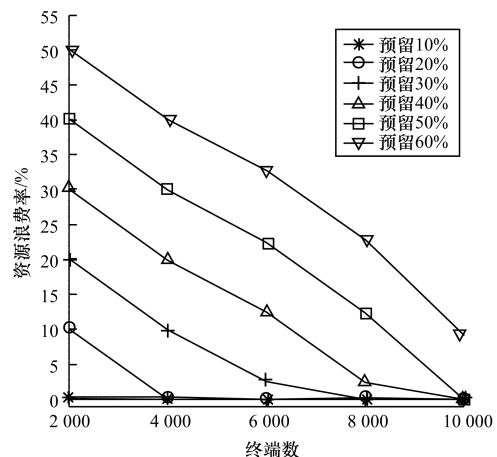


图 7 不同资源预留情况下的资源浪费率对比

本文通过业务驱动资源预留保证了高优先级终端的接入率,提高网络的总响应率,如图8所示,可以看出,在网络资源抢占机制中,高、低优先级终端的响应率和总响应率非常近似。

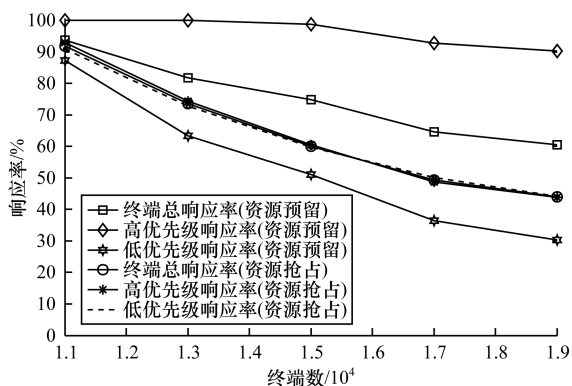
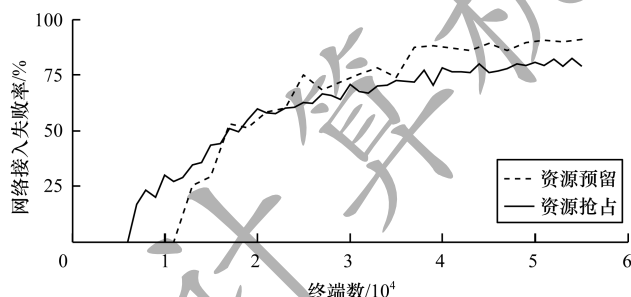
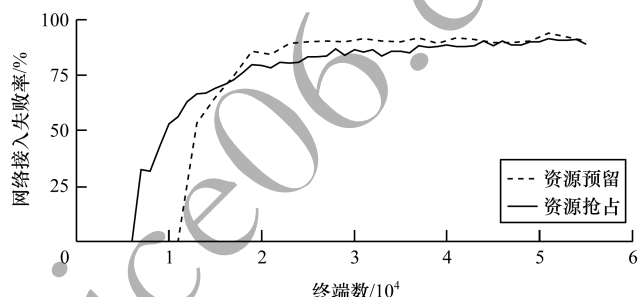


图8 资源预留和资源抢占机制下的终端响应率对比

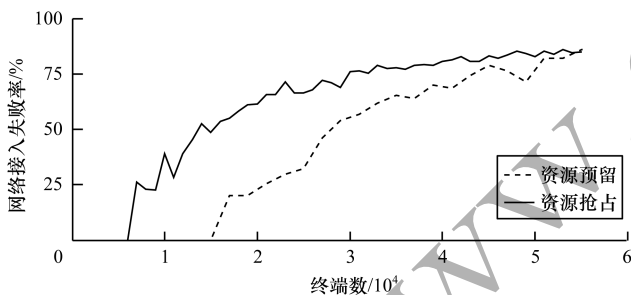
各优先级在不同机制下的网络接入失败率如图9所示,可以看出:每个终端以到达的先后顺序选择符合需求的最优网络接入,因此当终端量达到



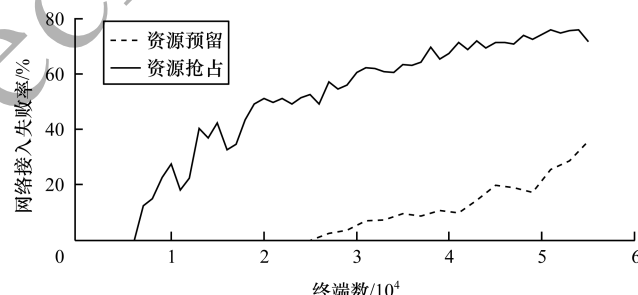
(a)优先级1



(b)优先级2



(c)优先级3



(d)优先级4

图9 各优先级在不同机制下的网络接入失败率对比

6 结束语

本文提出一种业务驱动资源预留的密集移动终端网络接入方法,用于热点高容量地区异构网络资源的分配,解决高优先级用户接入网络失败的问题。通过资源预留规划网络资源的使用情况,避免因网络资源紧缺导致终端请求网络服务失败的情况,从而减少网络接入时延。实验模拟大型商圈中密集移动终端的运动状态,结果表明,该方法可以根据终端

7 000时,所有的业务均出现一部分终端无法接入网络,而本文所提出的资源预留机制则最大限度地保障了高优先级终端的网络接入,提高了终端总响应率;当实验区域内终端数较少时,高优先级终端的响应率几乎达到100%,有效地保证了优先级分别为4和3的实时支付类业务和语音类业务的网络接入率;当终端数量达到30 000以上,支付类业务才开始出现接入失败。

在实验中暂将阈值 Th 的值设为90%,当实验区域内的终端数量达到20 000以上时,优先级为2的视频业务接入失败率达到90%以上。触发公平机制保证网络的公平性,高优先级接入率降低。终端密度过大时,如图9(a)和图9(b)所示,低优先级终端的接入失败率虽略比资源抢占机制高,但并不会出现终端持续等待接入网络的现象。由于每个网络内所有低优先级终端均轮流获得网络资源后,进入新一轮的资源预留,因此在密集高容量的场景下高优先级终端依然保持较高的接入率,如图9(c)和图9(d)所示。

的业务优先级和实际需求选择最接近理想的目标网络,通过资源预留的方式为高优先级终端预留网络资源,在尽可能少牺牲低优先级终端网络服务质量的同时,保障高优先级终端的用户体验质量,同时利用时间触发公平机制保障网络的公平性,避免低优先级终端出现“饿死”现象。下一步将研究如何提高低优先级终端的接入率,同时解决多终端并发切换情况下的网络资源分配问题。

参考文献

- [1] ONIRETI O, QADIR J, IMRAN M A, et al. Will 5G see its blind side? evolving 5G for universal Internet access [C]//Proceedings of GAIA '16. New York, USA: ACM Press, 2016: 1-6.
- [2] MAHINDRA R, VISWANATHAN H, SUNDARESAN K, et al. A practical traffic management system for integrated LTE-WiFi networks [C]//Proceedings of MobiCom '14. New York, USA: ACM Press, 2014: 189-200.
- [3] 蒋溢, 成哲, 邹洋. 基于多业务的移动性垂直切换算法[J]. 计算机工程, 2017, 43(1): 168-171, 177.
- [4] WONG W, THAKUR A, CHAN S H G. An approximation algorithm for AP association under user migration cost constraint [C]//Proceedings of IEEE INFOCOM '16. Washington D. C., USA: IEEE Press, 2016: 1-9.
- [5] 张振浩, 梁俊, 肖楠, 等. 空天异构网络中基于 Q 学习的切换判决优化算法[J]. 计算机工程, 2018, 44(5): 296-302, 308.
- [6] KIM Y, MOON K, LEE T J. Enhanced random access method with a coverage level and subcarrier spacing in NB-IoT systems [C]//Proceedings of ICIT '17. New York, USA: ACM Press, 2017: 292-296.
- [7] FABINI J, ZSEBY T, FABINI J, et al. The right time: reducing effective end-to-end delay in time-slotted packet-switched networks [J]. IEEE/ACM Transactions on Networking, 2016, 24(4): 2251-2263.
- [8] KWON Y M, SHAH S T, CHUNG M Y, et al. Performance analysis of small-cell base station with cellular and WiFi RATs [C]//Proceedings of IMCOM '15. New York, USA: ACM Press, 2015.
- [9] SUGIHARA S, HIROTA Y, FUJII S, et al. Dynamic resource allocation for immediate and advance reservation in space-division-multiplexing-based elastic optical networks [J]. IEEE/OSA Journal of Optical Communications and Networking, 2017, 9(3): 183-197.
- [10] OU Qinghai, ZHAO Honghao, YE Yuepian, et al. Dynamic resource allocation based on q-learning for VNE in fiber-wireless (FiWi) access network [C]//Proceedings of ICGSP '17. New York, USA: ACM Press, 2017: 116-120.
- [11] DELIMITROU C, KOZYRAKIS C. HCloud: resource-efficient provisioning in shared cloud systems [J]. ACM SIGOPS Operating Systems Review, 2016, 50(2): 473-488.
- [12] YEREMENKO O S, LEMESHKO O V, NEVZOROVA O S, et al. Method of hierarchical QoS routing based on network resource reservation [C]//Proceedings of the 1st IEEE Ukraine Conference on Electrical and Computer Engineering. Washington D. C., USA: IEEE Press, 2017: 971-976.
- [13] MAMMAN M, HANAPI Z M, ABDULLAH A, et al. An adaptive call admission control with bandwidth reservation for downlink LTE networks [J]. IEEE Access, 2017, 5: 10986-10994.
- [14] 陈赓, 夏玮玮, 沈连丰. 基于多门限预留机制的自适应带宽分配算法[J]. 通信学报, 2014, 35(12): 78-83.
- [15] SZYMANSKI T H. An ultra-low-latency guaranteed-rate Internet for cloud services [J]. IEEE/ACM Transactions on Networking, 2016, 24(1): 123-136.
- [16] YAO Chuting, GUO Jia, YANG Chenyang. Achieving high throughput with predictive resource allocation [C]//Proceedings of IEEE Global Conference on Signal and Information Processing. Washington D. C., USA: IEEE Press, 2016: 768-772.
- [17] FUERST C, SCHMID S, SURESH L, et al. Kraken: online and elastic resource reservations for cloud datacenters [J]. IEEE/ACM Transactions on Networking, 2018, 26(1): 422-435.
- [18] CONDOLUCI M, MAHMOODI T, STEINBACH E, et al. Soft resource reservation for low-delayed teleoperation over mobile networks [J]. IEEE Access, 2017, 5: 10445-10455.
- [19] NIU Di, XU Hong, LI Baochun. Resource auto-scaling and sparse content replication for video storage systems [J]. ACM Transactions on Modeling and Performance Evaluation of Computing Systems, 2017, 2(4): 1-19.
- [20] 陶洋, 刘小虎, 陶玲. 基于 QoS 的异构网络垂直切换算法[J]. 计算机工程, 2016, 42(12): 91-96.
- [21] 范存群, 王尚广, 孙其博, 等. 基于认知自选择决策树的垂直切换方法研究[J]. 通信学报, 2013, 34(11): 71-80.
- [22] 刘军, 李晓楠. 基于层次分析法的 WLAN/蜂窝网络切换判决算法[J]. 通信学报, 2013, 34(2): 65-72.

编辑 金胡考